## Searches for lepton number violation, and flavour violation beyond the Yukawa couplings at LHCb

Greg Ciezarek

High Energy Physics Blackett Laboratory Imperial College London

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy

 $20^{\rm th}$  August 2014

## Declaration

The work presented in this thesis was carried out between October 2010 and March 2014. It is the result of my own studies, with the support of members of the Imperial College HEP group and the broader LHCb collaboration. The work of others is explicitly referenced.

I made following specific contributions:

- 1. In Chapter 4 I developed the selection criteria and mass fit, measured the muon (mis)identification efficiencies in data, and studied the systematics related to these contributions.
- 2. In Chapter 5 I performed the entire analysis
- 3. In Chapter 6 I performed the entire study

This thesis was written by me, and has not been submitted for any other qualification. The work in Chapter 4 and 5 is published in Ref. [1] and Ref. [2], respectively.

Greg Ciezarek, 20 August 2014

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work (Imperial College London PhD regulations 2012).

All plots that are marked as coming from papers of the LHCb collaboration are available under the CC-BY 3.0 license, http://creativecommons.org/licenses/by/3.0/.

### Abstract

The Standard Model does not describe several phenomena, such as gravity and dark matter, and therefore is an incomplete description of nature. This demands the existence of new physics beyond the Standard Model. Two searches for new physics are presented in this thesis, along with a sensitivity study for a third analysis sensitive to new physics.

The  $\nu MSM$  model motivates a search for lepton number violation using  $B^+ \rightarrow h^- \mu^+ \mu^+$  decays, where  $h = (\pi, K)$ . No  $B^+ \rightarrow h^- \mu^+ \mu^+$  candidates are seen in  $\sim 36 \text{ pb}^{-1}$  of LHCb data and limits are set of  $\mathcal{B}(B^+ \rightarrow K^- \mu^+ \mu^+) < 4.1 \times 10^{-8}$  and  $\mathcal{B}(B^+ \rightarrow \pi^- \mu^+ \mu^+) < 4.4 \times 10^{-8}$  at 90% C.L. These improve the previous best limits by a factor 40 and 30, respectively.

Using ~ 1 fb<sup>-1</sup> of LHCb data, the  $B^+ \to \pi^+ \mu^+ \mu^-$  decay is observed for the first time with 5.2 $\sigma$  significance. This is the first  $b \to d\mu^+\mu^-$  transition to be observed. The  $B^+ \to \pi^+ \mu^+ \mu^-$  branching fraction is measured to be  $(2.3 \pm 0.6 \text{ (stat)} \pm 0.1 \text{ (syst)}) \times 10^{-8}$ . The ratio of branching fractions between  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  is measured to be 0.053  $\pm$  0.014 (stat)  $\pm$  0.001 (syst), and this is used to determine a value of the ratio of quark mixing matrix elements  $|V_{\rm td}|/|V_{\rm ts}| = 0.266 \pm 0.035 \text{ (stat)} \pm 0.003 \text{ (syst)}$ . All of these results are compatible with the Standard Model expectations.

Previous measurements of the ratio of  $B \to D^{(*)}\tau^+\nu$  and  $B \to D^{(*)}\mu^+\nu$  branching fractions exceed the Standard Model expectations by more than  $3\sigma$ , combining D and  $D^*$ . These decays are challenging to measure at a hadron collider, due to the presence of neutrinos in the final state. A sensitivity study is presented for a measurement of the ratio of  $B^0 \to D^{*-}\tau^+\nu$  and  $B^0 \to D^{*-}\mu^+\nu$  branching fractions at LHCb. This study includes a novel fit method, and two new algorithms which enable the backgrounds to be controlled, and control samples to be isolated. The estimated uncertainty on  $R_{D^*}$ , including the largest systematic uncertainties, is ~ 8%, competitive with the 9% uncertainty on the present best measurement of  $R_{D^*}$ .

## Acknowledgements

I am well aware of how lucky I have to be in the position to work on what I have, it has been a special time to enter the field. I owe a debt to the vast number of people who spent years developing and running the LHC accelerator and LHCb detector, and gave me such excellent data to play with. Throughout my time at Imperial and CERN, I have been surrounded by brilliant and creative people, and have appreciated this immensely.

I would like to thank my supervisor Mitesh Patel for everything he has done for me over the years. The time spent teaching me during the early stages of my time at Imperial was especially valuable. I also learned a huge amount from Tom Blake, at the cost of long periods of his time. Thanks also go to the rest of the Imperial LHCb group, in particular Patrick Owen, Indrek Sepp, Kostas Petridis, Ulrik Egede and Andrey Golutvin. There are many other people who made my time at Imperial and CERN enjoyable, especially all of the other students I spent so much time with.

Finally, I would like to thank my family, for being like a family to me.

# Contents

Li	List of Figures 8			
Li	List of Tables 15			
1	Intr	oducti	on	17
<b>2</b>	The	eory of	flavour violating interactions	19
	2.1	Flavou	r in the Standard Model	19
	2.2	Limita	ations of the Standard Model	23
	2.3	The fla	avour problem	24
		2.3.1	The Minimal Flavour Violation hypothesis	25
		2.3.2	The $\nu MSM$	25
	2.4	Theore	etical predictions of flavour changing hadron decays	26
		2.4.1	The Operator Product Expansion	26
		2.4.2	QCD effects in hadron decays	28
		2.4.3	Standard Model expectation for $\mathcal{B}(B^+ \to h^+ \mu^+ \mu^-)$	29
		2.4.4	Standard Model expectation for $\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu)$	31
3	Det	ector		35
	3.1	Heavy	flavour at the LHC	35
	3.2	Tracki	ng detectors and magnet	37
		3.2.1	The VELO	39
		3.2.2	The TT	39
		3.2.3	The IT	40
		3.2.4	The OT	40
	3.3	Calori	meters	41
	3.4	Charge	ed hadron identification	41
	3.5	The m	uon system	45
	3.6	Track	reconstruction	47
	3.7	Definit	tion of reconstructed quantities	48
		3.7.1	Reconstructed track quantities	48
		3.7.2	Reconstructed vertex quantities	48

3.8 Simulated events		ated events	49	
	3.9	Perfor	mance of the LHCb detector	49
	3.10	Trigge	я	53
		3.10.1	Hardware trigger	53
		3.10.2	The first stage of the High Level Trigger	55
		3.10.3	The second stage of the high level trigger	57
	3.11	Multiv	variate selection	59
4	Sea	rch for	the lepton number violating decay $B^+  ightarrow h^- \mu^+ \mu^+$	62
	4.1	Introd	uction	62
	4.2	Signal	selection and combinatoric background $\ \ldots \ $	63
		4.2.1	Control channels	66
		4.2.2	Combinatorial background estimation	66
	4.3	Peakir	ng backgrounds	66
		4.3.1	Mis-id probabilities	67
		4.3.2	Total misidentified background	68
	4.4	Norma	Alisation	70
	4.5	Limit	setting	71
	4.6	System	natic uncertainties	73
		4.6.1	Background expectations	73
		4.6.2	Single event sensitivity	73
	4.7	Result	s and conclusion	76
<b>5</b>	Firs	t obse	rvation of $B^+ \rightarrow \pi^+ \mu^+ \mu^-$	77
	5.1	Introd	uction	77
		5.1.1	Overview	78
	5.2	Event	selection $\ldots \ldots \ldots$	78
		5.2.1	Dimuon Mass Veto	79
		5.2.2	Combinatorial backgrounds	79
		5.2.3	Peaking and partially reconstructed backgrounds	83
		5.2.4	Control channels	84
	5.3	Miside	entified $B^+ \to K^+ \mu^+ \mu^-$ lineshape	85
	5.4	Signal	yield determination $\ldots$	85
		5.4.1	Reconstructed $B^+ \to J/\psi K^+$ candidates $\ldots \ldots \ldots \ldots \ldots \ldots$	87
		5.4.2	Reconstructed $B^+ \rightarrow J/\psi K^+$ candidates with the pion mass hy-	
			pothesis	87
		5.4.3	Reconstructed $B^+ \to \pi^+ \mu^+ \mu^-$ and $B^+ \to K^+ \mu^+ \mu^-$ candidates	88
		5.4.4	Cross check of the fit procedure $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	88
	5.5	Norma	alisation	90
		5.5.1	Reconstruction and selection efficiencies	90

		5.5.2	Change of form factors	. 9	1
	5.6	System	natics	. 9	2
		5.6.1	Fit uncertainties	. 9	2
		5.6.2	Normalisation uncertainties	. 9	3
		5.6.3	Normalisation to $B^+ \to K^+ \mu^+ \mu^-$	. 9	6
	5.7	Result	s and conclusion	. 9	6
6	Son	aitivity	study for $B^0 \setminus D^{*-} \tau^+ \mu$	0	Q
U	6 1	Introdu	uction	9 0	8
	6.2	Evneri	mental challenges	. 5	0
	6.3	Backo	rounds	. 10	1
	0.0	6.3.1	$B \rightarrow D^{**} \mu^+ \mu$ decays	. 10	1
		6.3.2	$B \to D^{(*)}DX$ decays	. 10	3
		6.3.3	Misidentified muon backgrounds	. 10	4
		6.3.4	Combinatorial background	. 10	4
	6.4	Signal	vield determination	. 10	4
		6.4.1	Selection criteria	. 10	6
		6.4.2	Fit to a toy dataset	. 11	0
	6.5	Backgi	round rejection $\ldots$	. 11	0
		6.5.1	Displaced $\tau$ decays	. 11	0
		6.5.2	Multivariate isolation	. 11	3
		6.5.3	Hadronic control channels	. 11	5
	6.6	Sensiti	vity to $R_{D^*}$	. 12	1
		6.6.1	Systematic uncertainties	. 12	2
	6.7	Conclu	$sion \ldots \ldots$	. 12	8
7	Con	clusior	18	12	9
Bi	Bibliography 132				
G	Glossary 141				
A	Appendix 142				
$\mathbf{A}$	A $B^+ \to \pi^+ \mu^+ \mu^-$ data-simulation comparisons 143				

# List of Figures

2.1	Allowed mass (M <sub>2</sub> ) and squared mixing angle to muon neutrinos ( $\theta_2^2$ ) for the heavier two sterile neutrinos [3]. The constraints are described in the	
	text	27
2.2	Example of a weak decay in the complete electroweak theory (left), and	
	effective theory (right)	27
3.1	Distributions of the angles between the $b$ quark and $\overline{b}$ quark momenta and	
	the beamline, as generated by PYTHIA 6	36
3.2	Schematic diagram of a side view of the LHCb detector, with subsystems	
	labelled	37
3.3	Schematic diagram of a VELO sensor.	38
3.4	Layout of the TT for an $x$ layer (a) and a $u$ layer (b)	39
3.5	Layout of the IT for an $x$ layer (a) and a $u$ layer (b)	40
3.6	Layout of the OT (a) and end view of an individual OT module (b)	40
3.7	Schematic diagram of the readout granularity of the ECAL (a) and HCAL	
	(b)	41
3.8	Polar angle versus momentum for pions from simulated $B^0 \to \pi^+\pi^-$ events.	42
3.9	Schematic diagram of RICH1 (a) and RICH2 (b)	43
3.10	Reconstructed Cherenkov angle in the $C_4F_{10}$ radiator versus momentum	
	for tracks with well isolated Cherenkov rings	44
3.11	Schematic diagram of the muon system.	45
3.12	Schematic diagram showing the different track types	47
3.13	The ratio of track reconstruction efficiency between 2011 data and simu-	
	lated events, as a function of momentum and pseudorapidity	50
3.14	IP resolution in measured in data (black) and simulation (red), as a function	
	of inverse $p_{\mathrm{T}}$	50
3.15	Comparisons between data and simulated $B^+ \rightarrow J/\psi K^+$ events for $B^+$	
	$\mathrm{IP}\chi^2.$ The unsmeared, shown in blue (medium grey), and smeared, shown	
	in light grey, simulated events are compared to data, shown in black. $\ . \ .$ .	51

3.16	The PID efficiency for kaons, and the pion misidentification rate, as a func-	
	tion of momentum. Efficiencies are shown for requirements of $DLL_{K\pi} > 0$	
	and $\text{DLL}_{K\pi} > 5$ , using $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$ decays selected from data	
	taken in 2012	51
3.17	Efficiency of the isMuon requirement for muons (a), protons (b), pions (c),	
	and kaons (d), as a function of momentum. Several ranges of $p_{\rm T}$ are shown.	52
3.18	Efficiency for L0Muon and L0DiMuon as a function of $J/\psi$ $p_{\rm T}$ for	
	$B^+ \rightarrow J/\psi K^+$	54
3.19	Efficiency for L0Hadron as a function of $B p_{\rm T}$ for $B^- \rightarrow D^0 \pi^-$ and	
	$B^0 \to D^-\pi^+$ , and as a function of $D p_{\rm T}$ for $D^0 \to K^-\pi^+$ and $D^+ \to K^-\pi^+\pi^+$ .	55
3.20	Efficiency for Hlt1TrackMuon as a function of $J/\psi p_{\rm T}$ for $B^+ \to J/\psi K^+$	56
3.21	Efficiency for Hlt1TrackAllL0 as a function of $B p_{\rm T}$ for $B^- \to D^0 \pi^-$ and	
	$B^0 \to D^-\pi^+$ , and as a function of $D p_{\rm T}$ for $D^0 \to K^-\pi^+$ and $D^+ \to K^-\pi^+\pi^+$ .	57
3.22	Efficiency for one of the Hlt2TopoBBDTnBody or	
	Hlt2TopoMuBBDTnBody $(n=2,3,4)$ algorithms as a function of B	
	$p_{\rm T}$ for $B^+ \to J/\psi K^+$ . The efficiency is measured relative to events which	
	have passed Hlt1TrackAllL0 or Hlt1TrackMuon.	58
3.23	Efficiency for one of the Hlt2TopoBBDTnBody (n=2,3,4) algorithms as	
	a function of $B p_{\rm T}$ for $B^- \to D^0 \pi^-$ and $B^0 \to D^- \pi^+$ . The efficiency is	
	measured relative to events which have passed Hlt1TrackAllL0	58
3.24	Schematic view of a decision tree	61
4.1	Generic s-channel Feynman diagram for $\Delta L = 2$ meson decays. The s-	
	channel makes the dominant contribution in processes with a sterile neu-	
	trino on the mass-shell [4]	63
4.2	The $M_{K^+\mu^+\mu^-}$ mass distribution of $K^+\mu^+\mu^-$ events after the application	
	of the selection criteria (a) selecting the $J/\psi$ invariant mass window, and	~
	(b) excluding the $J/\psi$ and $\psi(2S)$ invariant mass windows.	65
4.3	The $M_{\mu^+\mu^-}$ distribution of tag and probe events after the application of	
	the kinematic selection and trigger requirements: (a) with no additional	
	requirements, (b) requiring $DLL_{K\pi} > 1$ , (c) requiring isMuonLoose = 0,	~ ~
	and (d) requiring both $DLL_{K\pi} > 1$ and is MuonLoose = 0	69
4.4	The $M_{h^-\mu^+\mu^+}$ distributions for all of the misidentified backgrounds consid-	
	ered, for $B^+ \to K^- \mu^+ \mu^+$ (left) and $B^+ \to \pi^- \mu^+ \mu^+$ (right), scaled to their	
	respective expectations	70
4.5	The relative efficiency between $B^+ \to K^- \mu^+ \mu^+$ and $B^+ \to J/\psi K^+$ across	
	phase space.	72

4.6	The relative efficiency between the signal and normalisation channels as a function of $M_{h^-\mu^+}$ . The efficiencies at each value of $M_{h^-\mu^+}$ are calculated for each of the three extreme phase space distributions detailed in the text	79
4.7	The invariant mass distribution of $K^-\mu^+\mu^+$ (a) and $\pi^-\mu^+\mu^+$ (b) events after the application of the selection criteria. No events in either decay	12
	mode survive the selection criteria	75
4.8	The 90% C.L limits set on the branching fractions of the $B^+ \to K^- \mu^+ \mu^+$ (light-coloured line) and $B^+ \to \pi^- \mu^+ \mu^+$ (dark-coloured line) decays, as a	70
	$\text{function of } m_{h^-\mu^+}.$	70
5.1	The dimuon and $K^+\mu^+\mu^-$ mass for selected $K^+\mu^+\mu^-$ events, before appli- cation of any mass vetoes.	80
5.2	The dimuon mass veto illustrated with randomly generated mass points.	80
5.3	The dimuon and $K^+\mu^+\mu^-$ mass for selected $K^+\mu^+\mu^-$ events, after appli- cation of the dimuon mass veto.	80
5.4	Normalised distributions of $\Delta P(\mu^+\mu^-)$ for signal, shown in black, and side-	
	band, shown in blue (medium grey) regions of $B^+ \to J/\psi K^+$ data, and for	
	simulated $B^+ \to \pi^+ \mu^+ \mu^-$ events, shown in light grey)	81
5.5	Normalised distributions of the number of primary vertex degrees of free-	
	dom for signal, shown in black and sideband, shown in light grey, regions	
	of $B^+ \to J/\psi K^+$ data.	82
5.6	The significance obtained from toy fits, based on different BDT and $DLL_{\mu\pi}$	
	cuts	83
5.7	BDT output distribution for simulated $B^+ \to \pi^+ \mu^+ \mu^-$ events (black solid	
	line) and candidates taken from the mass sidebands in the data (red dotted	
	line)	84
5.8	The mass distribution of selected simulated $B^+ \to J/\psi K^+$ events, shown	
	in grey, and $B^+ \to K^+ \mu^+ \mu^-$ events, shown in black. In (b), $B^+ \to J/\psi K^+$	
	has been reweighted to match the $B^+\!\to K^+\mu^+\mu^-$ distributions in hadron	
	momentum and dimuon mass	86
5.9	Invariant mass distribution for $B^+ \to J/\psi K^+$ candidates under the (a)	
	$K^+\mu^+\mu^-$ and (b) $\pi^+\mu^+\mu^-$ mass hypotheses with the fit projections overlaid.	87
5.10	Invariant mass distribution of $B^+ \to \pi^+ \mu^+ \mu^-$ candidates with the fit pro-	
	jection overlaid (a) in the full mass range and (b) in the region around the	
	B mass	88
5.11	Invariant mass distribution of $B^+ \to K^+ \mu^+ \mu^-$ candidates with the fit pro-	
	jection overlaid (a) in the full mass range and (b) in the region around the	
	nominal $B$ mass	89

5.12	5.12 Invariant mass distribution of $B^+ \to J/\psi \pi^+$ candidates with the fit projec-		
	tion overlaid. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ 89		
5.13	Comparisons between data and simulated $B^+ \rightarrow J/\psi K^+$ events for the		
	BDT output. The smeared, shown in blue (medium grey) and unsmeared,		
	shown in light grey simulated events are compared to data, shown in black. 94		
5.14	Comparisons between the BDT output distributions for $B^+ \rightarrow J/\psi K^+$		
	candidates selected from the data, shown in black, and from simulated		
	events after reweighting, shown in grey		
6.1	Feynman diagram of $B \to D^{(*)} \tau^+ \nu$ , including a hypothetical charged Higgs		
	boson contribution.		
6.2	Existing measurements of $R_{\mathcal{D}(*)}$ ,, 99		
6.3	Predicted spectrum of $c\overline{u}$ bound states		
6.4	Distributions of true missing mass for simulated $B^0 \to D^{*-} \tau^+ \nu$ events.		
	shown in black, $B \to D^{**} \mu^+ \nu$ events, shown in light grey, and $B \to D^* D X$		
	events, shown in blue (darkest grey)		
6.5	Visible mass versus corrected mass for simulated $B^0 \to D^{*-} \mu^+ \nu$ events (top		
	left), $B^0 \to D^{*-} \tau^+ \nu$ events (top right), $B \to D^{**} \mu^+ \nu$ events (bottom left),		
	and $B \to D^*DX$ events (bottom left)		
6.6	Distributions of muon momentum (left) and transverse momentum (right)		
	for simulated $B^0 \to D^{*-} \mu^+ \nu$ events, shown in black, and $B^0 \to D^{*-} \tau^+ \nu$ ,		
	shown in purple (light grey)		
6.7	Distributions of $D^0$ momentum (left) and transverse momentum (right) for		
	simulated $B^0 \to D^{*-} \mu^+ \nu$ events, shown in black, and $B^0 \to D^{*-} \tau^+ \nu$ , shown		
	in purple (light grey)		
6.8	Visible mass versus corrected mass for simulated $B^0 \rightarrow D^{*-} \mu^+ \nu$ events (left		
	column) and $B^0 \to D^{*-} \tau^+ \nu$ events (right column), before imposing any		
	trigger requirements (top row), after requiring candidates to have passed		
	the topological trigger algorithms (middle row), and instead requiring can-		
	didates to have passed charm trigger algorithms (bottom row) 109		
6.9	Visible mass (left) and corrected mass (right) projections of a toy dataset. 110		
6.10	Schematic diagram of the decays $B^0 \to D^{*-} \mu^+ \nu$ (top), and $B^0 \to D^{*-} \tau^+ \nu$ ,		
	where the $\tau$ travels a greater distance than the $D^0$ before decaying (bottom).111		
6.11	Difference in the true decay positions between the $\tau$ and the $D^0$ for simu-		
	lated $B^0 \to D^{*-} \mu^+ \nu$ , shown in black, and $B^0 \to D^{*-} \tau^+ \nu$ , shown in purple		
	(light grey) events. Distances are shown in the direction parallel (left) and		
	perpendicular (right) to the beam direction		

6.12	Difference in the measured decay positions between the $\tau$ and the $D^0$ for	
	simulated $B^0 \to D^{*-} \mu^+ \nu$ , shown in black, and $B^0 \to D^{*-} \tau^+ \nu$ , shown in	
	purple (light grey) events. Distances are shown in the direction parallel	
	(left) and perpendicular (right) to the beam direction	112
6.13	Distribution of $D^0\mu$ vertex $\chi^2$ for $B^0 \to D^{*-}\mu^+\nu$ , shown in black, and	
	$B^0 \to D^{*-} \tau^+ \nu$ , shown in light grey	113
6.14	Distribution of flight MVA output for simulated $B^0 \rightarrow D^{*-} \mu^+ \nu$ (red	
	hatched) and $B^0 \to D^{*-} \mu^+ \nu$ (blue) events	114
6.15	Highest isolation MVA output for any track in simulated $B^0 \to D^{*-} \mu^+ \nu$	
	events (magenta) and simulated $B \rightarrow D^{**} \mu^+ \nu$ events (red hatched)	115
6.16	Comparison in performance between vertex isolation, shown in cyan (dark	
	grey) and the novel MVA isolation, shown in blue (light grey)	116
6.17	Distributions of $M_{D^{*-}\pi^+}$ , where the isolation MVA is used to add a pion to	
	$D^{*-}\mu^+$ candidates in data. Distributions are shown for track in the event	
	with the highest MVA output (top), and the next highest MVA output	
	(bottom). $\ldots$	117
6.18	The $M_{D^{*-}\pi^+\pi^-\pi^+}$ distribution of $B^0 \to D^{*-}\pi^+\pi^-\pi^+$ candidates	118
6.19	Isolation BDT output distributions for $B^0 \to D^{*-} \pi^+ \pi^- \pi^+$ candidates taken	
	from background-subtracted data, shown in black, and simulated events,	
	shown in purple (light grey)	119
6.20	Schematic diagrams of the $B^0 \rightarrow D^{*-}\pi^+\pi^-\pi^+$ decay, reconstructed as	
	$[D^{*-}\pi^+]$ (top), and $[K^-H^+]$ (bottom).	119
6.21	Flight BDT output distributions for $B^0 \to D^{*-}\pi^+\pi^-\pi^+$ candidates taken	
	from background-subtracted data, shown in black, and simulated events,	
	shown in purple (light grey). The flight BDT is evaluated using the $[D^{*-}\pi^+]$	
	vertex, mimicking $B^0 \to D^{*-} \mu^+ \nu$ (left), and the $[K^- H^+]$ vertex, mimicking	
	$B^0 \to D^{*-} \tau^+ \nu$ (right).	120
6.22	Visible mass (left) and corrected mass (right) projections of a toy dataset	
	in region 0I. This region requires candidates to be isolated, and have no	
	measured tau flight	122
6.23	Visible mass (left) and corrected mass (right) projections of a toy dataset	
	in region 0A. This region requires candidates to be anti-isolated, and have	
	no measured tau flight	123
6.24	Visible mass (left) and corrected mass (right) projections of a toy dataset in	
	region 1I. This region requires candidates to be isolated, and have moderate	
	measured tau flight	123
6.25	Visible mass (left) and corrected mass (right) projections of a toy dataset	
	in region 1A. This region requires candidates to be anti-isolated, and have	
	moderate measured tau flight	123

6.26	Visible mass (left) and corrected mass (right) projections of a toy dataset	
	in region 2I. This region requires candidates to be isolated, and have high	
	measured tau flight	124
6.27	Visible mass (left) and corrected mass (right) projections of a toy dataset	
	in region 2A. This region requires candidates to be anti-isolated, and have	
	high measured tau flight	124
6.28	Distribution of $M_{D^{*-}\pi^+}$ for $D^{*+}\pi^-\mu^-\nu$ candidates	125
6.29	Distribution of $M_{D^{*-}\pi^+\pi^-}$ for $D^{*-}\pi^+\pi^-\mu^+\nu$ candidates.	125
6.30	Distribution of $R_{D^*}$ measured from a toy dataset fitted with different values	
	used for the $B \to D^{**} \mu^+ \nu$ form factors	127
71	Measured value of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ angular observable $P'$ as a function	
1.1	of dimuon mass	130
79	Batios between experimental measurements and lattice OCD SM expecta-	100
1.2	tions for exclusive $h \rightarrow s \mu^+ \mu^-$ decays at high dimuon mass	131
	tions for exclusive $b \neq s\mu^{\mu}$ decays at high diffuon mass	101
A.1	Comparisons between data and simulated $B^+ \to J/\psi  K^+$ events for the $B^+$	
	$\chi^2_{IP}.$ The smeared (red) and unsmeared (green) simulated events are shown	
	against data (black).	143
A.2	Comparisons between data and simulated $B^+ \to J\!/\psiK^+$ events for relevant	
	$B^+$ variables in $B^+ \to J/\psi K^+$	144
A.3	Comparisons between data and simulated $B^+ \to J/\psi K^+$ events for relevant	
	$K^+$ variables in $B^+ \to J/\psi K^+$	144
A.4	Comparisons between data and simulated $B^+ \to J\!/\psiK^+$ events for relevant	
	$\mu^+$ variables in $B^+ \to J/\psi K^+$	145
A.5	Comparisons between data and simulated $B^+ \to J\!/\psiK^+$ events for relevant	
	$\mu^-$ variables in $B^+ \to J/\psi K^+$	145
A.6	Comparisons between data and simulated $B^+ \rightarrow J/\psi K^+$ events for	
	$\Delta P(\mu^+\mu^-)$ in $B^+ \to J/\psi K^+$ .	146
A.7	Comparisons between data and simulated $B^+ \to J/\psi K^+$ events for the $B^+$	
	$\chi^2_{IP}$ , after reweighting	146
A.8	Comparisons between data and simulated $B^+ \to J/\psi K^+$ events for relevant	
	$B^+$ variables in $B^+ \to J/\psi K^+$ , after reweighting	146
A.9	Comparisons between data and simulated $B^+ \to J/\psi K^+$ events for relevant	
	$K^+$ variables in $B^+ \to J/\psi K^+$ , after reweighting	147
A.10	Comparisons between data and simulated $B^+ \to J/\psi K^+$ events for relevant	
	$\mu^+$ variables in $B^+ \to J/\psi K^+$ , after reweighting	147
A.11	Comparisons between data and simulated $B^+ \to J/\psi K^+$ events for relevant	
	$\mu^-$ variables in $B^+ \to J/\psi K^+$ , after reweighting	148

# List of Tables

2.1	Constraints on new physics scales from various neutral meson mixing tran- sitions, assuming coupling constants of one [5]	25
3.1	LHC running conditions in 2010–2012, together with the amount of inte-	
	grated luminosity recorded by the LHCb detector	36
3.2	The isMuon identification requirements.	46
3.3	The isMuonLoose identification requirements.	46
4.1	The selection criteria used in the preselection.	64
4.2	The selection criteria.	65
4.3	A summary of misidentified backgrounds for $B^+ \to K^- \mu^+ \mu^+$ .	70
4.4	A summary of misidentified backgrounds for $B^+ \to \pi^- \mu^+ \mu^+$ .	71
4.5	Sources of systematic error and their relative uncertainty on the single	
	event sensitivity for $B^+ \to K^- \mu^+ \mu^+$ and $B^+ \to \pi^- \mu^+ \mu^+$ .	75
51	The selection criteria used in the preselection	70
5.9	The variables used in the BDT	79 81
5.2	Values for the efficiencies shown in Eqn. (5.5), calculated using simulated	01
	events.	91
5.4	Values for the efficiencies shown in Eqn. (5.5) for $B^+ \to \pi^+ \mu^+ \mu^-$ and	
	$B^+ \to K^+ \mu^+ \mu^-$ , calculated using simulated events.	91
5.5	Summary of normalisation systematics	96
5.6	Summary of systematics in normalising to $B^+ \to K^+ \mu^+ \mu^-$ .	96
61	Measured masses and widths of $D^{**}$ states with various quantum numbers	102
6.2	HEAG averages of $\mathcal{B}(B \to D^{**}\ell^+\nu) \times \mathcal{B}(D^{**} \to D^{(*)}\pi^+)$ measurements [6]	102
6.3	Measured $B^+ \rightarrow X_c \mu^+ \nu$ inclusive and exclusive decay modes, corrected for	10-
0.0	$D^{**}$ decay modes [7].	102
6.4	Measured $D^+ \rightarrow e^+ X$ inclusive and exclusive decay modes [8]. Assuming	
	the $\overline{K}^0 \pi^0 e^+ \nu_e$ final state is 50% the rate of $K^- \pi^+ e^+ \nu_e$ (as is the case for	
	the dominant $K^*(892)^0$ resonance), the measured exclusive decay modes	
	saturate the inclusive rate.	103

6.5	Measured $D^0 \to e^+ X$ inclusive and exclusive decay modes [8]. The mea-	
	sured exclusive decay modes saturate the inclusive rate	
6.6 Measured $D_s^+ \to e^+ X$ inclusive and exclusive decay modes [8]. The r		
	sured exclusive decay modes saturate the inclusive rate	
6.7	List of PDFs included in the fit, along with the source of the events used	
	to produce the template	
6.8	The selection criteria used in the preselection	
6.9	Selection requirements applied after the preselection	
6.10	Input variables for the flight BDT	
6.11	Input variables for the isolation MVA	
6.12	Definition of the binning scheme in flight BDT output and isolation BDT	
	output	
6.13	Additional requirements applied to select the $D^{*-}\pi^+\mu^+\nu$ final state. $~$ 123	
6.14	Additional requirements applied to select the $D^{*-}\pi^+\pi^-\mu^+\nu$ final state. $\ . \ . \ 124$	

## Chapter 1

### Introduction

Following the discovery of the Higgs boson, all of the fundamental particles predicted by the Standard Model (SM) of particle physics have been observed. The SM is a selfconsistent theory, with many predictions verified to a high level of precision. Such predictions include the existence, mass and width of the  $Z^0$  boson, where the latter two predictions have been confirmed to sub-percent accuracy. However, the SM does not accommodate all phenomena observed in nature. A particularly glaring shortcoming of the SM is that it is unable to account for  $\sim 95\%$  of the energy in universe, with the unexplained portion thought to consist of dark matter and dark energy. Gravity, one of the two forces which most shapes our day-to-day existence, as well as shaping the universe on the largest scales, is not included in the SM. While the 5% of the universe consisting of baryonic matter is described by the SM, the means by which this matter came into existence is not; the Big Bang should have produced an equal amount of matter and antimatter, which would have annihilated to produce a universe containing only light. Subsequent interactions must have changed the proportions of matter and antimatter, resulting in the matter dominated universe we see today. The rate of such interactions required to produce the observed matter-antimatter asymmetry is  $\sim 10$  orders of magnitude greater than that predicted by the SM. All of these problems demand the existence of 'new physics' beyond the SM.

Another curious feature of the SM is the existence of three 'copies', known as generations, of each type of fermion. The electron, for example, has two copies in the form of the muon and the tau, differing only in mass. Similarly, the up quark and down quark each have two heavier copies, giving a total of six types ('flavours') of quark. The only known differences between the three generations arise due to the Yukawa couplings between the Higgs field and the fermions; the Yukawa couplings generate the fermion masses, as well as all of the interactions which transform between the six flavours of quark.

In addition to the mystery of the three generations, there are other problems with the origins of the SM parameters, the vast majority of which also involve the Yukawa couplings. This therefore motivates quark flavour changing decays as a place to search for the effect of new physics. Flavour changing processes are sensitive to new physics via the virtual contributions of new particles. Historically, a number of major discoveries in particle physics have been inferred from virtual effects before direct observations were made: the charm quark was predicted to explain the approximately eight orders of magnitude between the rates of the  $K_L^0 \to \mu^+\mu^-$  and  $K^+ \to \mu^+\nu$  decays<sup>1</sup>; the third generation of quarks was predicted due to the observation of charge-parity (CP) violation in kaon decays; the high mass of the top quark was then predicted due to the observation of *B* meson mixing, and subsequently by precision measurements of electroweak processes.

The LHCb experiment is dedicated to making precision measurements of flavour violating processes. It is located at the Large Hadron Collider (LHC), in order to exploit the large number of hadrons produced containing heavy flavoured quarks. The LHCb detector is described in Chapter 3, along with the techniques used to reconstruct and select decays.

This thesis presents two analyses of LHCb data, and a sensitivity study for a future LHCb analysis. The theoretical techniques used to compute predictions for flavour violating decays are introduced in Chapter 2, along with the  $\nu MSM$  model. This model motivates a search for lepton number violation using  $B^+ \rightarrow h^- \mu^+ \mu^+$  decays, where  $h = (\pi, K)$ , which is presented in Chapter 4. The second analysis is a search for flavour violation beyond the Yukawa couplings using  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  and  $B^+ \rightarrow K^+ \mu^+ \mu^-$  decays, which is presented in Chapter 5. A sensitivity study for a test of generation universality using  $B^0 \rightarrow D^{*-} \tau^+ \nu$  and  $B^0 \rightarrow D^{*-} \mu^+ \nu$  decays is presented in Chapter 6. Finally, conclusions are presented in Chapter 7.

<sup>&</sup>lt;sup>1</sup>Charge conjugation is implicit throughout this thesis.

### Chapter 2

# Theory of flavour violating interactions

The SM contains two forces: the electroweak force, which gives rise to all of the flavour changing interactions in the SM, and the strong force. The mechanism of flavour changing interations in the electroweak sector is described in Sect. 2.1. The strong force, described by Quantum ChromoDynamics (QCD) is not discussed in any detail in this chapter, as it does not give any flavour changing interactions. The most important feature of QCD is that at low energies the theory is non-perturbative, and therefore calculations cannot be made.

The limitations of the SM, and the motivation for searching for new physics in flavour violating interactions are discussed in Sect. 2.2. The constraints on new physics from flavour violating processes are discussed in Sect. 2.3. A general means of circumventing these constraints, the Minimal Flavour Violation hypothesis, is discussed. Alternatively, a specific model, the Minimal Neutrino Standard Model ( $\nu MSM$ ), which circumvents these constraints, and which motivates a search for  $B^+ \to h^-\mu^+\mu^+$  decays, is also introduced. The effective theory framework with which flavour violating meson decays are calculated is discussed in Sect. 2.4, including techniques for calculating hadronic form factors. The applications of these theoretical techniques to calculate the SM expectations for  $B^+ \to h^+\mu^+\mu^-$  and  $B^0 \to D^{*-}\ell^+\nu$  are then presented.

### 2.1 Flavour in the Standard Model

The quark flavour changing interactions in the SM are determined entirely by the interactions of the quarks with the Higgs field<sup>1</sup>. The Higgs mechanism is introduced to break the electroweak symmetry, generating mass terms for the  $W^{\pm}$  and  $Z^{0}$  bosons while avoiding introducing a photon mass term. In addition, the quarks and charged leptons all gain

<sup>&</sup>lt;sup>1</sup>This section is written with reference to Ref. [9]

their masses via 'Yukawa couplings' to the Higgs field and, in the SM it is these Yukawa couplings which are source of all flavour violating interactions. In the SM, a single complex Higgs doublet is introduced:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.1}$$

where  $\phi^+$  ( $\phi^0$ ) are electromagnetically charged (electromagnetically neutral) complex scalar fields. The Higgs Lagrangian is

$$\mathcal{L}_{\text{Higgs}} = -(D_{\mu}\phi)^{\dagger}(D^{\mu}\phi) - V(\phi^{\dagger}\phi) + \mathcal{L}_{\text{Y}}, \qquad (2.2)$$

where the first term is a kinetic term,  $V(\phi^{\dagger}\phi)$  is the Higgs potential, and  $\mathcal{L}_{Y}$  is a Lagrangian containing the Yukawa interactions, described below. The Higgs potential is given by

$$V(\phi^{\dagger}\phi) = \lambda \left(\phi^{\dagger}\phi\right)^{2} - \mu^{2}\phi^{\dagger}\phi + \frac{\mu^{4}}{4\lambda},$$
(2.3)

where  $\mu$  is the Higgs mass parameter and  $\lambda$  is the Higgs self coupling. For  $\mu^2 > 0$ , the minimum of this potential is at  $\phi = 0$ , a point around which the electroweak Lagrangian is symmetric. However, for values of  $\mu^2 < 0$  the minimum is at

$$|\phi| = \sqrt{\frac{\mu^2}{2\lambda}},\tag{2.4}$$

a point around which the electroweak Lagrangian is no longer symmetric. If the Higgs potential is initially set to a value of  $\phi = 0$ , the electroweak symmetry is still intact. However, this point is not stable against quantum fluctuations, and so  $|\phi|$  takes an expectation value in a vacuum of

$$\langle |\phi| \rangle = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{1}{\sqrt{2}}\nu, \qquad (2.5)$$

spontaneously breaking the electroweak symmetry. The parameter  $\nu$  is referred to as the vacuum expectation value, and is measured to be  $\nu = 246 \text{ GeV}$  [10]. This defines the energy scale at which the electroweak symmetry is broken, referred to as the 'electroweak scale'. In the 'unitary gauge', the direction of the vacuum expectation value is chosen to be

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0\\ \nu \end{pmatrix}. \tag{2.6}$$

Excitations around this minimum produce a neutral scalar field

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0\\ (\nu + H) \end{pmatrix}, \qquad (2.7)$$

corresponding to the physical Higgs boson.

The Yukawa interaction term for a charged lepton is

$$\mathcal{L}_{\mathbf{Y}}^{\ell} = -g_{\ell} \left( \bar{\ell}_L \phi \ell_R + \bar{\ell}_R \phi^{\dagger} \ell_L \right), \qquad (2.8)$$

where  $g_e$  is the Yukawa coupling strength for the lepton,  $\ell_{L,R}$  are the left- and right-handed components of the lepton field and  $\phi$  is the Higgs doublet. In the unitary gauge, this can be rewritten as

$$\mathcal{L}_{\mathbf{Y}}^{\ell} = -\frac{1}{\sqrt{2}} g_{\ell} \nu (\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L), \qquad (2.9)$$

$$= -\frac{1}{\sqrt{2}}g_{\ell}\nu\bar{\ell}\ell. \tag{2.10}$$

which corresponds to a lepton mass term with a value of

$$m_{\ell} = \frac{1}{\sqrt{2}} g_{\ell} \nu. \tag{2.11}$$

Similarly, the Yukawa interaction term for the quarks is

$$\mathcal{L}_{\mathbf{Y}}^{q} = \left(g_{ij}^{u}\bar{q}_{Li}\phi_{C}u_{Rj} + g_{ij}^{d}\bar{q}_{Li}\phi d_{Rj} + \text{h.c.}\right), \qquad (2.12)$$

where the i and j indicies run over the three generations, and h.c. refers to the hermitian conjugate of the expression given. The left-handed and right-handed quark matrix doublets have the following structure:

$$q_{Li} = \begin{pmatrix} u_{Li} \\ d_{Li} \end{pmatrix}, \qquad u_{Ri} = \begin{pmatrix} u_{Ri} \\ 0 \end{pmatrix}, \qquad d_{Ri} = \begin{pmatrix} 0 \\ d_{Ri} \end{pmatrix}.$$
(2.13)

As with Eqn. 2.10, the quark Yukawa Lagrangian can be written in terms of mass matrices by working in the unitary gauge

$$\mathcal{L}_{Y}^{q} = -\bar{u}_{Li}m_{ij}^{u}u_{Rj} - \bar{d}_{Li}m_{ij}^{d}d_{Rj} + h.c., \qquad (2.14)$$

where  $m_{ij}^u$  and  $m_{ij}^d$  are the mass matrices for up and down type quarks respectively. These matrices are constructed in the generational basis. In general, these matrices have off-diagonal terms, and therefore do not directly represent the mass basis. To move into the mass basis, the mass matrices are diagonalised using four unitary matrices  $V_{L,R}^{u,d}$ :

$$M_{diag}^{u} = V_{L}^{u} m_{u} V_{L}^{u\dagger} \qquad \qquad M_{diag}^{d} = V_{L}^{d} m_{d} V_{L}^{d\dagger}. \qquad (2.15)$$

The diagonalised mass matrices take the form

$$m_{\alpha}^{u} = \begin{pmatrix} m_{u} & 0 & 0 \\ 0 & m_{c} & 0 \\ 0 & 0 & m_{t} \end{pmatrix}, \qquad m_{\alpha}^{d} = \begin{pmatrix} m_{d} & 0 & 0 \\ 0 & m_{s} & 0 \\ 0 & 0 & m_{b} \end{pmatrix}.$$
(2.16)

To produce the physical quark masses, the strengths of the quark Yukawa couplings must be:

The origin of the values of these parameters is not currently known.

In the generational basis, the interactions between the  $W^{\pm}$  boson and the quarks are given by the following Lagrangian

$$\mathcal{L}_{\rm CC} = \frac{ig_2}{\sqrt{2}} \left[ W^+_{\mu} \bar{u}_{Lj} \gamma^{\mu} d_{Lj} + W^-_{\mu} \bar{d}_{Lj} \gamma^{\mu} u_{Lj} \right].$$
(2.18)

To transform into the mass basis, the four diagonalisation matrices are inserted as follows:

$$\mathcal{L}_{\rm CC} = \frac{ig_2}{\sqrt{2}} \left[ W^+_{\mu} \bar{u}_{L\alpha} \left[ V^u_L V^{d\dagger}_L \right] \gamma^{\mu} d_{L\beta} + W^-_{\mu} \bar{d}_{L\alpha} \left[ (V^d_L V^{u\dagger}_L \right] \gamma^{\mu} u_{L\beta} \right].$$
(2.19)

The product  $V_{\text{CKM}} \equiv V_L^u V_L^{d\dagger}$  is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix [11]. This matrix transforms between the generation eigenstates (i.e. the quark flavours) and the mass eigenstates

$$\begin{pmatrix} d'\\s'\\b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub}\\V_{cd} & V_{cs} & V_{cb}\\V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d\\s\\b \end{pmatrix} = V_{\rm CKM} \begin{pmatrix} d\\s\\b \end{pmatrix}$$
(2.20)

where d, s, b represent the flavour eigenstates and d', s', b' represent the mass eigenstates. If the off-diagonal elements of the CKM matrix are non-zero, as is observed in nature, then the weak current can couple between generations, and therefore can change the quark flavour. The CKM matrix is commonly expressed in the Wolfenstein parameterisation

$$V_{\rm CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \qquad (2.21)$$

where

$$\lambda = \frac{|V_{us}|}{\sqrt{|V_{us}|^2 + |V_{ud}|^2}} \qquad A\lambda^2 = \left|\frac{V_{us}}{V_{us}}\right| \qquad A\lambda^3(\rho + i\eta) = V_{ub}^*.$$
(2.22)

The value of the  $\lambda$  parameter is measured to be  $\lambda = 0.2252 \pm 0.0009$  [8], so the CKM matrix displays a strong hierarchy between diagonal terms, and those furthest off-diagonal.

Similarly to the charged-current case, the interaction between the  $Z^0$  boson and the quarks in the generational basis is given by

$$\mathcal{L}_{\rm NC} = i \sqrt{g_1^2 + g_1^2} \bigg[ Z_\mu \bar{u}_j \gamma^\mu \left( g_V + \gamma_5 g_A \right) u_j + Z_\mu \bar{d}_j \gamma^\mu \left( g_V + \gamma_5 g_A \right) d_j \bigg], \qquad (2.23)$$

where  $g_V$  and  $g_A$  are the vector and axial-vector charges respectively. Transforming into the mass basis requires the insertion of diagonalisation matrix products of the type  $V^q V^{q\dagger}$ 

$$\mathcal{L}_{\rm NC} = i \sqrt{g_1^2 + g_1^2} \bigg[ Z_\mu \bar{u}_j \gamma^\mu \left[ V^u V^{u\dagger} \right] \left( g_V + \gamma_5 g_A \right) u_j + Z_\mu \bar{d}_j \gamma^\mu \left[ V^d V^{d\dagger} \right] \left( g_V + \gamma_5 g_A \right) d_j \bigg].$$
(2.24)

By the unitarity of the diagonalisation matrices, the products  $V^{u}V^{u\dagger}$  and  $V^{d}V^{d\dagger}$  are both identity, and the neutral current therefore conserves flavour.

### 2.2 Limitations of the Standard Model

The SM is known to be an incomplete theory of nature. There are several observed phenomena which are not accounted for, such as the nature of dark matter and dark energy; the matter-antimatter asymmetry necessary to go from a net baryon number of zero at the big bang to the baryonic universe we see today; the mechanism which generates neutrino mass; and how to include gravity in the model.

In addition to the unexplained phenomena listed above, there are a number of other problems within the SM, all of which involve the Yukawa couplings. In these cases, rather than new phenomena, the problems are with the free parameters in the model. The SM has 18 free parameters, 13 of which lie in the Yukawa sector. At present, the mass of the charm quark is considered to be as fundamental a parameter as the strength of the strong interaction, or the weak mixing angle. In addition to the number of parameters, there is also the pattern of values the parameters take. The electron and top quark masses differ by a factor of  $10^6$ , with no apparent explanation for this difference. As clearly shown in the Wolfenstein parameterisation, the CKM matrix has a distinct hierarchy, with large on-diagonal elements and progressively smaller off-diagonal elements. This is suggestive of an underlying structure which generates the values of the Yukawa coupling parameters but presently the origin of this structure is unknown. The introduction of quark masses opens a problem in QCD; a CP violating phase is present, but experimentally this phase is limited by neutron electric dipole moments to be smaller than  $\sim 10^{-9}$ . As this parameter could in principle take any value, it being so close to zero is suggestive that there may be some new physics enforcing this requirement. This situation is avoided if the lightest

quark is massless, however the same argument applies with the relevant Yukawa coupling then being close to zero.

The only thing which renders the three generations non-degenerate are their Yukawa couplings, suggesting that whatever new physics determines the values of the Yukawa couplings is also the origin of the three-generation structure we observe.

A final problem introduced by the Yukawa couplings is a quadratically divergent Higgs mass. In the absence of other interactions, the Higgs mass is determined entirely by the Higgs self-coupling. The physical Higgs mass, however, is also shifted by interactions between the Higgs and the gauge bosons, and the fermions. The size of the correction from the fermions grows quadratically with increasing energy, and at large energies a considerable cancellation between these corrections and the Higgs self-interaction term is required to obtain the light Higgs mass that is observed.

The scale at which any of these problems are solved is in general unclear; the only scale readily identifiable with any of these problems is the Planck scale, where gravity becomes strongly coupled. Due to the apparent weakness of gravity, this scale is  $\mathcal{O}(10^{19})$  GeV, a far higher energy than experimentally accessible in the foreseeable future. The central question of particle physics is whether there is any new physics present at a scale significantly below the Planck scale.

### 2.3 The flavour problem

In general, new physics models introduce new sources of flavour violation. Flavour changing processes therefore set bounds on the energy scale and couplings of such models. The constraints set on the energy scale and coupling are degenerate; a bound corresponds to either a higher energy scale with a stronger coupling, or a lower energy scale with a weaker coupling. In order to keep such constraints as general as possible, a 'natural' coupling constant of unity is usually considered.

In Ref. [5], measurements made prior to LHC running are used to set constraints on the scale of new physics. These constraints are summarised in Tab. 2.1. The processes considered are  $K^0 - \overline{K}^0$ ,  $D^0 - \overline{D}^0$ ,  $B_d^0 - \overline{B_d^0}$  and  $B_s^0 - \overline{B_s^0}$  mixing. These processes are ordered by the extent to which they are suppressed in the SM. The bounds in Tab. 2.1 assume no new CP-violating phases are present, the constraints are stronger still if such phases are introduced. If new physics models are assumed to violate flavour equally for all three generations, the strongest constraints come from  $K^0 - \overline{K}^0$  mixing. These agreement between measurements and the SM predictions set an energy scale for new physics of  $\gtrsim 10^4$  TeV. The weakest constraints are from assuming new physics contributes only to  $B_s^0 - \overline{B_s^0}$  mixing, but even in this case the energy scale of new physics is  $\gtrsim 10^2$  TeV, far above the LHC energy. If any new physics is to be seen in direct production at the LHC, it must therefore either have very small flavour violating couplings, or not violate flavour at

Transition	Bound on $\Lambda$ (TeV)
$(\overline{s}_L \gamma^u d_L)^2$	$9.8 \times 10^2$
$(\overline{s}_R d_L)(\overline{s}_L d_R)$	$1.8 \times 10^4$
$(\overline{c}_L \gamma^u u_L)^2$	$1.2 \times 10^3$
$(\overline{c}_R u_L)(\overline{c}_L u_R)$	$1.8 \times 10^4$
$(\overline{b}_L \gamma^u d_L)^2$	$5.1 \times 10^2$
$(\overline{b}_R d_L)(\overline{b}_L d_R)$	$1.9 \times 10^3$
$(\overline{b}_L \gamma^u s_L)^2$	$1.1 \times 10^2$
$(\overline{b}_R s_L)(\overline{b}_L s_R)$	$3.7  imes 10^2$

Table 2.1: Constraints on new physics scales from various neutral meson mixing transitions, assuming coupling constants of one [5].

all. As new physics is widely expected at the  $\sim$  TeV scale, in order to cancel the quadratic divergence in the Higgs mass, this is known as the 'flavour problem'.

#### 2.3.1 The Minimal Flavour Violation hypothesis

The Minimal Flavour Violation (MFV) hypothesis is a means of circumventing the flavour problem, by stating that the only sources of flavour violation, even in new physics models, are the Yukawa couplings. This ensures that the strengths of all flavour changing interactions are governed by the CKM matrix, and that no new flavour changing neutral current interactions are introduced. Deviations from the SM expectations are still possible under the MFV hypothesis but processes which in the SM are related by different CKM matrix elements are forced to retain the same correlation in any new physics scenario.

By forcing new physics models to obey MFV, the constraints in Tab. 2.1 are considerably weakened, allowing the possibility of new physics being seen via direct production at the LHC. Searches for supersymmetry at the LHC now assume MFV as a matter of course. The necessity of imposing MFV is therefore one of the most important experimental constraints on new physics models at the TeV scale. However, it is important to test this hypothesis, by measuring ratios of observables which are related by MFV. In particular, constructing tests of MFV using loop decays, which are sensitive to new physics significantly above the TeV scale, may allow the observation of non-MFV new physics.

#### 2.3.2 The $\nu MSM$

Another means of circumventing the flavour problem is for new physics to introduce only new particles which couple extremely weakly with the SM particles. One such theory is the Neutrino Minimal Standard Model ( $\nu MSM$ ), described in Ref. [12]. This model is intended to be the minimal addition to the SM which accommodates neutrino oscillations, baryogenesis, and dark matter, while remaining consistent with cosmological data. This therefore accounts for three of the phenomena left unexplained in the SM listed in Sect. 2.2, leaving only gravity and dark energy. In this model, both of these problems are solved by other new physics introduced at the Planck scale, and, in order to be consistent with experimental observations, no other new physics is therefore required below the Planck scale [13].

The  $\nu MSM$  introduces three massive gauge singlet neutral leptons, with both Dirac and Majorana mass terms. The lightest of these is proposed as a dark matter candidate, while the heavier two give rise to baryogenesis. In order to comply with astrophysical observations, the dark matter candidate is constrained to have a mass of order 10 keV/ $c^2$ . To produce the baryon asymmetry in the universe, the heavier two neutrinos are required to be near degenerate, with a mass in the range between 0.15 and 100  $\text{GeV}/c^2$  [14]. In order for the two heavier sterile neutrinos to maintain the element abundances produced by primordial nucleosynthesis, they must decay on a timescale shorter than approximately 0.1 s, and this introduces lower bounds on their mixing [15]. The heavier two sterile neutrinos would enhance the branching fraction of processes such as  $B^+ \rightarrow h^- \mu^+ \mu^+$  [4]. The branching fraction of such processes is proportional to the fourth power of the active-sterile neutrino mixing angle, as the sterile neutrino must mix with the SM neutrino twice: once to produce the sterile neutrino, and again in order for it to decay to SM particles. The region in the mixing angle – majorana neutrino mass plane that is allowed by current experimental measurements is shown in Fig. 2.1 [3]. The upper limit in mixing angle (labelled "BAU" on Fig. 2.1) comes from the requirement that the sterile neutrinos do not come into thermal equilibrium too soon and hence destroy any baryon asymmetry produced. The lower limit in mixing angle and mass labelled "BBN" on Fig. 2.1 comes from the requirement that the sterile neutrino lifetime is short enough to preserve the predictions for Big Bang nucleosynthesis. The other lower limit in mixing angle, labelled "See-Saw", comes from the requirement that the masses generated for the active neutrinos are sufficiently large to be consistent with neutrino oscillation data.

A search for the  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$  decays, which are potentially sensitive to the heavier two majorana neutrinos, is presented in Chapter 4. This search is sensitive in the region of sterile neutrino mass accessible on-shell in *B* decays.

### 2.4 Theoretical predictions of flavour changing hadron decays

#### 2.4.1 The Operator Product Expansion

Flavour changing quark decays contain both QCD and electroweak contributions. These two classes of contribution are characterised by very different time and distance scales, and so are largely decoupled from one another. This separation may be used to construct an effective theory, which incorporates all of our theoretical knowledge up to a certain energy



Figure 2.1: Allowed mass  $(M_2)$  and squared mixing angle to muon neutrinos  $(\theta_2^2)$  for the heavier two sterile neutrinos [3]. The constraints are described in the text.



Figure 2.2: Example of a weak decay in the complete electroweak theory (left), and effective theory (right).

scale, and parameterises our ignorance of physics far above that scale. The canonical example of such a theory is the Fermi theory of the weak interactions [16]. In this theory, weak interactions are reduced to an effective four-fermion vertex, as demonstrated in Fig. 2.2, with a coupling

$$G_F = \frac{\sqrt{2}}{8} \frac{g^2}{m_W^2} = 1.16637 \times 10^{-5} \,\text{GeV}/c^2.$$
(2.25)

This relation demonstrates that the full theory allows the calculation of the coupling constant in the effective theory.

The Operator Product Expansion (OPE) is an effective theory which separates low energy (QCD) contributions from high energy (electroweak) contributions. In this case, the purpose is to test for possible contributions from new physics at higher energies. The OPE defines an effective Hamiltonian of the form

$$\mathcal{H}_{\text{eff}} = \sum_{i} \mathcal{C}_{i} \mathcal{O}_{i}, \qquad (2.26)$$

where the 'operators',  $\mathcal{O}_i$ , define a particular effective process *i* (e.g. the effective fourfermion interaction of the Fermi theory). The 'Wilson coefficients',  $\mathcal{C}_i$ , determine the coupling of the effective interaction, as with  $G_F$ . These Wilson coefficients include the SM electroweak contributions, which are calculable, and any potential new physics contributions. By comparing the measured Wilson coefficients to the SM expectations, new physics can therefore be searched for in a model independent way. Conversely, the contributions to the Wilson coefficients made by a specific new physics model may be calculated, allowing that model to be tested against the data.

The SM contributions to the Wilson coefficients are calculated at a scale  $\mu$  characterising the weak interaction,  $\mu = m_W$ . For *B* decays the more relevant scale is  $\mu = m_b$ , and so the Wilson coefficients are 'evolved' down to  $\mu = m_b$ . As the scale above which effects are 'integrated out' has been reduced, effects which were once encoded in the long distance operators are instead transferred into the Wilson coefficients for different operators. This effect is known as 'operator mixing'. Wilson coefficients which incorporate the mixing in of other operators are denoted  $C_i^{\text{eff}}$  [17].

The operators and calculated Wilson coefficients can then be used to construct a matrix element, by inserting the operators between an initial and final state. For exclusive final states, this requires the introduction of hadronic form factors.

#### 2.4.2 QCD effects in hadron decays

Flavour changing hadron decays suffer from non-perturbative QCD contributions, the calculation of which is problematic. In particular, exclusive decays require the transition from a quark level process to a particular final state hadron to be calculated. This hadronisation process is described in terms of hadronic form factors, which are specific to a given initial and final state. A great deal of theoretical work has been devoted to calculating such hadronic form factors, and methods now exist for calculating form factors for processes involving a single final state hadron. Examples of such methods include Heavy Quark Effective Theory (HQET), Light Cone Sum Rules (LCSR), and lattice QCD, each of which are described below.

The best understood form factors in B meson decays are those for  $B \to D$  transitions. In this case, HQET produces reliable results. Due to their large masses relative to the energy scale of QCD,  $\Lambda_{\rm QCD}$ , the 'heavy' (b or c) quarks may be treated as having an infinite mass and, expanding around this limit, an effective theory containing additional symmetries may be constructed [18]. A heavy quark in a meson may therefore be treated as a stationary point source of colour charge, analogous to a proton providing a point source of electric charge in a hydrogen atom. The light quark is treated as interacting with the colour potential induced by the heavy quark, rather than interacting with it directly. Form factors in HQET may be calculated under the premise that a heavy quark decaying into another heavy quark does not change the potential which interacts with the light quark. Corrections to the heavy quark. This works particularly well when the velocity transferred in the b quark decay is small i.e. the stationary point source remains to a good approximation stationary. Calculations in the HQET work very well for  $b \to c$  transitions, for example  $B \to D^{(*)} \ell^+ \nu$  decays, as both the initial and final quarks are sufficiently heavy.

However, transitions of a *b* quark to lighter quarks, e.g.  $b \to s$  or  $b \to d$  transitions are not well described by HQET, as the mass of the final state quark is below  $\Lambda_{\text{QCD}}$ . Additional methods must therefore be employed to calculate such form factors. One such method is that of LCSR [19,20]. In this case the limit considered is that of large hadron momentum (large energy limit). An expansion is made around the large energy limit in powers of the distance between partons in the direction transverse to the hadron momentum [21,22].

While non-perturbative QCD cannot be solved analytically, lattice QCD techniques can offer a numerical solution. Lattice calculations of both heavy-to-heavy and heavyto-light form factors have been made, offering a greater precision than other techniques over certain kinematic ranges. The numerical techniques do not deal well with significant meson velocities, and lattice techniques are therefore best suited to the region of lowest hadronic recoil, as is the case for HQET.

### 2.4.3 Standard Model expectation for ${\cal B}(B^+ o h^+ \mu^+ \mu^-)$

As detailed in Sect. 2.3.1, it is important to test the MFV hypothesis by measuring ratios of observables. One such means of testing the MFV hypothesis is measuring the ratio between exclusive  $b \to s\mu^+\mu^-$  and  $b \to d\mu^+\mu^-$  decays. A measurement of this ratio is presented in Chapter 5. The effective Hamiltonian for the inclusive  $b \to q\mu^+\mu^-$  decay, where q = (s, d), takes the form<sup>2</sup>

$$\mathcal{H}_{\text{eff}} = -4 \frac{G_F}{\sqrt{2}} V_{tq}^* V_{tb} \sum_{i=1}^{10} \mathcal{C}_i(\mu) \mathcal{O}_i(\mu) , \qquad (2.27)$$

neglecting helicity suppressed contributions. The operators most relevant to the shortdistance contributions are  $\mathcal{O}_{7,9,10}$ 

$$\mathcal{O}_{7} = \frac{e}{g^{2}} m_{b}(\bar{q}_{L}\sigma^{\mu\nu}b_{R})F_{\mu\nu},$$
  

$$\mathcal{O}_{9} = \frac{e^{2}}{g^{2}}(\bar{q}_{L}\gamma_{\mu}b_{L})(\bar{\ell}\gamma^{\mu}\ell),$$
  

$$\mathcal{O}_{10} = \frac{e^{2}}{g^{2}}(\bar{q}_{L}\gamma_{\mu}b_{L})(\bar{\ell}\gamma^{\mu}\gamma_{5}\ell),$$
(2.28)

where e is the electro-magnetic coupling strength, g is the weak coupling strength,  $\sigma_{\mu\nu}$  are the Pauli spin matrices, and  $F^{\mu\nu}$  the electromagnetic field tensor. Ignoring interactions

<sup>&</sup>lt;sup>2</sup>This calculation is taken from Ref. [23]

between the b quark and the spectator quark, the effective Hamiltonian gives an amplitude

$$\mathcal{M}(b \to q\ell^+\ell^-) = \frac{G_F \alpha}{\sqrt{2}\pi} V_{tq}^* V_{tb} \left\{ \mathcal{C}_9^{\text{eff}} \left[ \bar{s} \gamma_\mu P_L b \right] \left[ \bar{\ell} \gamma^\mu \ell \right] + \mathcal{C}_{10} \left[ \bar{s} \gamma_\mu P_L b \right] \left[ \bar{\ell} \gamma^\mu \gamma_5 \ell \right] \right. \\ \left. -2 \hat{m}_b \mathcal{C}_9^{\text{eff}} \left[ \bar{s} i \sigma_{\mu\nu} \frac{\hat{q}^\nu}{\hat{s}} P_R b \right] \left[ \bar{\ell} \gamma^\mu \ell \right] \right\} .$$

where  $p_{\pm}$  is the four momentum of the  $\ell^{\pm}$ , q is the four-momentum of the dimuon system,  $s = q^2$ ,  $\hat{m}_b = m_b/m_B$ , and  $\hat{s} = s/m_B^2$ , where  $m_b$  and  $m_B$  are the masses of the b quark and B hadron, respectively. However, what is measured is not a free quark decay, but the decay of a hadron containing the initial quark, to a hadronic system containing the final quark. Calculating the amplitude of such a decay requires the non-perturbative QCD interactions within this initial and final hadronic system. The measurements presented in Chapter 5 will focus on the simplest choices of initial final hadron,  $B^+ \to h^+ \mu^+ \mu^-$ , where  $h = \pi, K$ . For  $B^+ \to h^+ \mu^+ \mu^-$ , the QCD contributions can be parameterised in terms of three form factors, defined by

$$\langle h(p)|\bar{s}\gamma_{\mu}b|B(p_B)\rangle = f_{+}(s)\left\{(p_B+p)_{\mu} - \frac{m_B^2 - m_h^2}{s}(p_B-p)_{\mu}\right\} + \frac{m_B^2 - m_h^2}{s}f_0(s)q_{\mu},$$
(2.29)

and

$$\langle h(p) | \bar{s} \sigma_{\mu\nu} q^{\nu} (1+\gamma_5) b | B(p_B) \rangle \equiv \langle h(p) | \bar{s} \sigma_{\mu\nu} (p_B - p)^{\nu} b | B(p_B) \rangle$$

$$= i \left\{ (p_B + p)_{\mu} s - (p_B - p)_{\mu} (m_B^2 - m_h^2) \right\}$$

$$\times \frac{f_T(s)}{m_B + m_h}.$$

$$(2.30)$$

Inserting these form factors into the amplitude for  $b \to q\mu^+\mu^-$  in Eqn. 2.29 gives the amplitude for  $B^+ \to h^+\mu^+\mu^-$ 

$$\mathcal{M} = \frac{G_F \alpha}{2\sqrt{2\pi}} V_{tq}^* V_{tb} m_B \left[ \left( A'(\hat{s}) \, \hat{p}_\mu + B'(\hat{s}) \, \hat{Q}_\mu \right) \left( \bar{\ell} \, \gamma^\mu \, \ell \right) + \left( C'(\hat{s}) \, \hat{p}_\mu + D'(\hat{s}) \, \hat{Q}_\mu \right) \, \left( \bar{\ell} \, \gamma^\mu \, \gamma_5 \, \ell \right) \right]$$

$$\tag{2.31}$$

with

$$A'(\hat{s}) = C_9^{\text{eff}}(\hat{s}) f_+(\hat{s}) + \frac{2\hat{m}_b}{1+\hat{m}_h} C_7^{\text{eff}} f_T(\hat{s}) , \qquad (2.32)$$

$$B'(\hat{s}) = \mathcal{C}_{9}^{\text{eff}}(\hat{s}) f_{-}(\hat{s}) - \frac{2\hat{m}_{b}}{\hat{s}}(1-\hat{m}_{h})\mathcal{C}_{7}^{\text{eff}}f_{T}(\hat{s}) , \qquad (2.33)$$

$$C'(\hat{s}) = C_{10} f_+(\hat{s}) ,$$
 (2.34)

$$D'(\hat{s}) = C_{10} f_{-}(\hat{s}) ,$$
 (2.35)

where  $\hat{m}_h = m_h/m_B$  and  $\hat{Q} = (P_B - p)/m_B$ . The rate for  $B^+ \to h^+ \mu^+ \mu^-$  as a function of

 $\hat{s}$  is then given by

$$\begin{aligned} \frac{\mathrm{d}\Gamma_h}{\mathrm{d}\hat{s}} &= \frac{G_F^2 \alpha^2 m_B^5}{2^{10} \pi^5} \left| V_{tq}^* V_{tb} \right|^2 \hat{u}(\hat{s}) \times \left\{ (|A'|^2 + |C'|^2) (\lambda - \frac{\hat{u}(\hat{s})^2}{3}) \right. \\ & \times |C'|^2 4 \hat{m}_\mu^2 (2 + 2\hat{m}_h^2 - \hat{s}) + \operatorname{Re}(C'D'^*) 8 \hat{m}_\mu^2 (1 - \hat{m}_h^2) + |D'|^2 4 \hat{m}_\mu^2 \hat{s} \right\} \end{aligned}$$

with

$$\hat{u}(\hat{s}) = \sqrt{\lambda(1 - 4\frac{\hat{m}_{\mu}^2}{\hat{s}})},$$
(2.36)

$$\lambda \equiv \lambda(1, \hat{m}_h^2, \hat{s}) = 1 + \hat{m}_h^4 + \hat{s}^2 - 2\hat{s} - 2\hat{m}_h^2(1 + \hat{s}) , \qquad (2.37)$$

where  $\hat{m}_{\mu} = m_{\mu}/m_B$ . The total width is given by integrating Eqn. 2.39

$$\Gamma_{h} = \frac{G_{F}^{2} \alpha^{2} m_{B}^{5}}{2^{10} \pi^{5}} \left| V_{tq}^{*} V_{tb} \right|^{2} \int_{\hat{s} = (2\hat{m}_{\mu})^{2}}^{\hat{s} = (1-\hat{m}_{h})^{2}} \hat{u}(\hat{s}) \times \left\{ (|A'|^{2} + |C'|^{2}) (\lambda - \frac{\hat{u}(\hat{s})^{2}}{3}) + |C'|^{2} 4\hat{m}_{\mu}^{2} (2 + 2\hat{m}_{h}^{2} - \hat{s}) + Re(C'D'^{*}) 8\hat{m}_{\mu}^{2} (1 - \hat{m}_{h}^{2}) + |D'|^{2} 4\hat{m}_{\mu}^{2} \hat{s} \right\} d\hat{s} \quad (2.38)$$

$$\equiv \left| V_{tq}^{*} V_{tb} \right|^{2} R_{h} \quad (2.39)$$

This gives an SM branching fraction for  $B^+ \to \pi^+ \mu^+ \mu^-$  of  $(2.0 \pm 0.2) \times 10^{-8}$  [24], with the dominant uncertainty coming from the calculation of the hadronic form factors. The SM expectation for the ratio of  $\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-)$  and  $\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)$  is then given by

$$\frac{\Gamma_{\pi}}{\Gamma_{K}} = \frac{|V_{td}^{*}V_{tb}|^{2}}{|V_{ts}^{*}V_{tb}|^{2}} \frac{R_{\pi}}{R_{K}}$$
(2.40)

$$\equiv \frac{|V_{td}|^2}{|V_{ts}|^2} R , \qquad (2.41)$$

where

$$R = \frac{R_{\pi}}{R_K}.$$
(2.42)

The measurement of the ratio of  $\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-)$  and  $\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)$  presented in Chapter 5, together with the calculated value of R therefore gives a measurement of  $|V_{\rm td}|/|V_{\rm ts}|$ . The SM expectation for R is discussed in detail in Sect. 5.7. If this value of  $|V_{\rm td}|/|V_{\rm ts}|$  deviates from that measured in other processes, particularly the very precise determination from  $B_d^0$  and  $B_s^0$  mixing, it would be evidence of non-MFV new physics.

### 2.4.4 Standard Model expectation for $\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu)$

In the SM, the decays  $B^0 \to D^{*-} \ell^+ \nu$  ( $\ell = \mu, \tau$ ) differ only due to the mass of the lepton. The differential rate of the  $B^0 \to D^{*-} \ell^+ \nu$  decay as a function of the decay kinematics is given by<sup>3</sup>

$$\frac{d^{2}\Gamma_{\ell}}{dq^{2}d\cos\theta} = \frac{G_{F}^{2}|V_{cb}|^{2}|p|q^{2}}{256\pi^{3}m_{B}^{2}} \left(1 - \frac{m_{\ell}^{2}}{q^{2}}\right)^{2} \times \left[(1 - \cos\theta)^{2}|H_{+}(q^{2})|^{2} + (1 + \cos\theta)^{2}|H_{-}(q^{2})|^{2} + 2\sin^{2}\theta|H_{0}(q^{2})|^{2} + \frac{m_{\ell}^{2}}{q^{2}}\left((\sin^{2}\theta(|H_{+}(q^{2})|^{2} + |H_{-}(q^{2})|^{2}) + 2|H_{s}(q^{2}) - H_{0}(q^{2})\cos\theta|^{2}\right)\right],$$

$$(2.43)$$

where  $q^2$  is the mass of the  $\ell\nu$  system,  $\theta$  is the angle between the  $D^*$  and the  $\ell$  in the  $\ell\nu$  rest frame. The  $H_X$  are the hadronic currents, which in the SM are given by

$$H_{\pm}^{\rm SM}(q^2) = (m_B + m_{D^*})A_1(q^2) \mp \frac{2m_B}{m_B + m_{D^*}} |p| V(q^2) ,$$
  

$$H_0^{\rm SM}(q^2) = \frac{1}{2m_{D^*}\sqrt{q^2}} \left[ (m_B^2 - m_{D^*}^2 - q^2)(m_B + m_{D^*})A_1(q^2) - \frac{4m_B^2|p|^2}{m_B + m_{D^*}}A_2(q^2) \right] ,$$
  

$$H_s^{\rm SM}(q^2) = \frac{2m_B|p|}{\sqrt{q^2}} A_0(q^2) .$$
(2.44)

where  $A_0(q^2), A_1(q^2), A_2(q^2)$  and  $V(q^2)$  are hadronic form factors defined by

$$\langle D^{*}(p_{D^{*}}, \epsilon_{\alpha}) | \bar{c}\gamma_{\mu} b | B(p_{B}) \rangle = \frac{2iV(q^{2})}{m_{B} + m_{D^{*}}} \epsilon_{\mu\nu\alpha\beta} \epsilon^{*\nu} p_{B}^{\alpha} p_{D^{*}}^{\beta},$$
(2.45a)  

$$A_{0}(q^{2}) \langle D^{*}(p_{D^{*}}, \epsilon_{\alpha}) | \bar{c}\gamma_{\mu}\gamma_{5}b | B(p_{B}) \rangle = 2m_{D^{*}} A_{0}(q^{2}) \frac{\epsilon^{*} \cdot q}{q^{2}} q_{\mu} + (m_{B} + m_{D^{*}}) A_{1}(q^{2})$$

$$\times \left( \epsilon_{\mu}^{*} - \frac{\epsilon^{*} \cdot q}{q^{2}} q_{\mu} \right)$$

$$- A_{2}(q^{2}) \frac{\epsilon^{*} \cdot q}{m_{B} + m_{D^{*}}}$$

$$\times \left( (p_{B} + p_{D^{*}})_{\mu} - \frac{m_{B}^{2} - m_{D^{*}}^{2}}{q^{2}} q_{\mu} \right).$$
(2.45b)

Integrating Eqn. 2.43 over  $\theta$  gives the rate of  $B \to D^{(*)} \ell^+ \nu$  as a function of  $q^2$ 

$$\frac{d\Gamma_{\ell}}{dq^2} = \frac{G_F^2 |V_{cb}|^2 |p| q^2}{96\pi^3 m_B^2} \left(1 - \frac{m_\ell^2}{q^2}\right)^2 \left[ \left(|H_+(q^2)|^2 + |H_-(q^2)|^2 + |H_0(q^2)|^2\right) \times \left(1 + \frac{m_\ell^2}{2q^2}\right) + \frac{3}{2} \frac{m_\ell^2}{q^2} |H_s(q^2)|^2 \right].$$
(2.46)

<sup>&</sup>lt;sup>3</sup>This calculation is taken from Ref. [25]

The ratio of  $B^0 \to D^{*-} \tau^+ \nu$  to  $B^0 \to D^{*-} \mu^+ \nu$  as a function of  $q^2$  is then given by

$$R_{D^*}(q^2) = \frac{d\Gamma_{\tau}/dq^2}{d\Gamma_{\ell}/dq^2} = \left(1 - \frac{m_{\tau}^2}{q^2}\right)^2 \left[\left(1 + \frac{m_{\tau}^2}{2q^2}\right) + \frac{3}{2}\frac{m_{\tau}^2}{q^2}\frac{|H_s(q^2)|^2}{|H_+(q^2)|^2 + |H_-(q^2)|^2 + |H_0(q^2)|^2}\right].$$
(2.47)

At a given point in the phase space, the only difference between  $B^0 \to D^{*-}\tau^+\nu$  and  $B^0 \to D^{*-}\mu^+\nu$  is given by the helicity suppressed form factor terms. Averaging Eqn. 2.47 across phase space results in a lower value of  $R_{D^*}$ , due to the considerably reduced phase space available to  $B^0 \to D^{*-}\tau^+\nu$ .

The form factors are calculable using the HQET technique. Defining the velocity transfer

$$w \equiv v_B \cdot v_{D^*} = \frac{m_B^2 + m_{D^*}^2 - q^2}{2m_B m_{D^*}},$$
(2.48)

the process is described by a single universal form factor

$$h_{A_1}(w) = A_1(q^2) \frac{1}{M} \frac{2}{w+1}, \qquad (2.49)$$

and the ratios  $R_1$ ,  $R_2$  and  $R_0$ 

$$A_{0}(q^{2}) = \frac{R_{0}(w)}{M} h_{A_{1}}(w) ,$$
  

$$A_{2}(q^{2}) = \frac{R_{2}(w)}{M} h_{A_{1}}(w) ,$$
  

$$V(q^{2}) = \frac{R_{1}(w)}{M} h_{A_{1}}(w) ,$$
(2.50)

where  $M = 2\sqrt{m_B m_{D^*}}/(m_B + m_{D^*})$ . In the heavy quark limit, the variation of these form factors with w is given by

$$h_{A_1}(w) = h_{A_1}(1) \left[ 1 - 8\rho^2 z + (53\rho^2 - 15)z^2 - (231\rho^2 - 91)z^3 \right],$$
  

$$R_1(w) = R_1(1) - 0.12(w - 1) + 0.05(w - 1)^2,$$
  

$$R_2(w) = R_2(1) + 0.11(w - 1) - 0.06(w - 1)^2,$$
  

$$R_0(w) = R_0(1) - 0.11(w - 1) + 0.01(w - 1)^2,$$
  
(2.51)

where  $z = (\sqrt{w+1} - \sqrt{2})/(\sqrt{w+1} + \sqrt{2})$ . The factor of  $h_{A_1}(1)$  is common to all of the hadronic currents in Eqn. 2.47, and  $R_{D^*}$  does not therefore depend on the value of  $h_{A_1}(1)$ . Three of the four remaining parameters have been measured using  $B^0 \to D^{*-}\ell^+\nu$  decays, with averages [6]:

$$\rho^2 = 1.401 \pm 0.033$$
  $R_1(1) = 1.401 \pm 0.033$   $R_2(1) = 0.854 \pm 0.020.$  (2.52)

The  $R_0(1)$  parameter has not been measured, as  $B^0 \to D^{*-}\mu^+\nu$  decays offer minimal sensitivity to this parameter due to the helicity suppression. This parameter must therefore be taken from theoretical calculations, such as those using HQET [25]. The resulting SM expectation for  $R_{D^*}$  is  $0.252 \pm 0.003$ , with the dominant uncertainty arising from the calculation of the  $R_0(1)$  parameter. The most precise measurement of this quantity is  $R_{D^*} = 0.332 \pm 0.024 \pm 0.018$ , from the BaBar collaboration [26]. This measurement, along with a simultaneous measurement or  $R_D$ , is more than three  $\sigma$  from the SM expectation. This strongly motivates a measurement of  $R_{D^*}$  with improved precision. A sensitivity study for a measurement of  $R_{D^*}$  is presented in Chapter 6.

### Chapter 3

### Detector

The LHC produces both the highest energy and the greatest rate of hadron-hadron interactions ever achieved at a particle collider. These two factors together result in an unprecedented rate of  $b\bar{b}$  and  $c\bar{c}$  production. The phase space available for the  $b\bar{b}$  to hadronise is sufficient to allow the production of  $B_s^0$  mesons and B baryons, which have not been produced in large quantities at previous dedicated B physics experiments. To take advantage of this large sample of B and D hadrons, the LHCb experiment is located on one of the eight LHC interaction points. The design aim of the LHCb experiment is to search for new physics in CP violation, or in rare B decays.

#### 3.1 Heavy flavour at the LHC

The LHC is a superconducting proton-proton collider with a circumference of 27km, spanning the French-Swiss border near Geneva. At nominal running conditions, it collides 7 TeV beams at a frequency of 40MHz. To date, the machine has operated at a maximum energy of 4 TeV per beam, at a maximum frequency of 20MHz. The LHCb detector cannot operate at the highest instantaneous luminosities achieved by the LHC during 2011 and 2012, as the detector occupancy becomes too high for acceptable performance. For this reason the LHC beams are not focussed as strongly at the LHCb interaction point as for the other experiments, resulting in a lower instantaneous luminosity. Additionally the two beams have a small offset at the collision point, reducing the overlap and so further reducing the luminosity. As the beams spend longer in the LHC the number of protons in the beams decreases, resulting in a decreasing luminosity with operating time. This may be countered by gradually reducing the offset between the two beams to compensate, allowing operation at a constant luminosity. This mode of operation, known as luminosity levelling, has been pioneered for LHCb operation in 2011 and 2012. The majority of the 2011 and 2012 dataset was collected at an instantaneous luminosity of  $4 \times 10^{32}$  cm<sup>-2</sup>s<sup>-1</sup>. The LHC running conditions for 2010–2012 are summarised in Tab. 3.1, along with the integrated luminosity recorded by the LHCb experiment.

Year	Beam energy (TeV)	Bunch crossing frequency (MHz)	Recorded luminosity $(fb^{-1})$
2010	3.5	20	0.04
2011	3.5	20	1
2012	4	20	2

Table 3.1: LHC running conditions in 2010–2012, together with the amount of integrated luminosity recorded by the LHCb detector. The analyses presented in Chapter 4 and 5 use solely the 2010 and 2011 datasets, respectively.



Figure 3.1: Distributions of the angles between the b quark and  $\overline{b}$  quark momenta and the beamline, as generated by PYTHIA 6.

The production of b quarks occurs predominantly at low angles to the beamline, as shown in Fig. 3.1. The LHCb detector is therefore designed as a single arm spectrometer. It consists of a number of subsystems, which are illustrated in Fig. 3.2. Each of these subsystems is described in Sects. 3.2–3.5. One of the distinguishing characteristics of Bmesons is their relatively long lifetime of ~ 1.5ps. Given their average boost, B mesons produced at the LHC travel a distance of  $\mathcal{O}(1 \text{ cm})$  before decaying. This characteristic makes precision measurements of the trajectory of charged particles as close to the interaction region as possible critical for B physics. This requirement drives the design of the VErtex LOcator (VELO) subdetector, described in Sect. 3.2, together with design of the remainder of the tracking system. The other key requirement of the tracking system is a good mass resolution, in order to separate a given B decay from background processes, and from the decays of other B hadrons.

Distinguishing different decays of B mesons requires the identity of individual particles to be determined. In particular, many decays, e.g the  $B^+ \to K^+ \mu^+ \mu^-$  and  $B^+ \to \pi^+ \mu^+ \mu^$ decays measured in Chapter 5, require kaons to be distinguished from pions. In order


Figure 3.2: Schematic diagram of a side view of the LHCb detector, with subsystems labelled. Figure taken from [27].

to identify hadrons a Ring Imaging CHerenkov (RICH) system is used. This system is described in Sect. 3.4. Leptons are identified using a dedicated muon system in the case of muons (described in Sect. 3.5), and a Calorimeter (described in Sect. 3.3) in the case of electrons.

In design running conditions, the LHC bunch crossing rate is 40 MHz. Given an event size of  $\mathcal{O}(100 \text{ kB})$ , storing every event would mean writing  $\mathcal{O}(\text{TB/s})$  to offline storage, which is unfeasible. In order to avoid this, while still taking advantage of the high interaction rate, data are stored in temporary buffers and a fast decision taken on which events are interesting and should be stored. This system is referred to as the trigger, and is described in Sect. 3.10. A  $b\bar{b}$  pair is present in only around one in every three hundred visible proton-proton interactions at  $\sqrt{s} = 7 \text{ TeV}$ , meaning that the trigger has to be highly selective. The ability to select *B* decays in the trigger algorithms, particularly for fully hadronic final states, was therefore a further crucial design criterion for the LHCb detector.

# 3.2 Tracking detectors and magnet

The tracking system has two purposes: a precise measurement of the initial trajectory of every charged particle, and of their deflection when traversing a strong magnetic field. The



Figure 3.3: Schematic diagram of a VELO sensor. Figure taken from [27].

production directions of charged particles enables the reconstruction of the proton-proton interaction vertices, referred to as Primary Vertex (PV), and any potential secondary vertices (SVs) arising from the decay of long-lived particles. The deflection of a charged particle in a magnetic field provides a measurement of the charge and momentum of the particle.

To this end, a dipole magnet providing a field with an integrated strength of ~ 4 Tm is employed, along with a sequence of three tracking detectors: the VErtex LOcator (VELO), followed by the Tracker Turicensis (TT), and the three tracking stations downstream of the magnet (T1-T3, collectively T stations). The T stations employ two different technologies, due to the large difference in occupancy between the inner and outer regions: an Inner Tracker (IT) consisting of silicon microstrips, and an Outer Tracker (OT) consisting of straw drift-tubes. The total baseline of the tracking system is ~10m, ensuring a large deflection distance for a given angular deflection.



Figure 3.4: Layout of the TT for an x layer (a) and a u layer (b). Dimensions are shown in cm. Figure taken from [28].

#### 3.2.1 The VELO

A schematic diagram of the VELO is shown in Fig. 3.3. The VELO consists of 21 modules, each comprising two layers of silicon microstrip sensor orientated perpendicular to the beamline. The two sensors in each module have strips aligned in the radial (r-sensor) and azimuthal ( $\phi$ -sensor) directions. For both sensors, the strip pitch varies with proximity to the beamline, from 40 µm in the innermost region to 100 µm at the outer edge. To give a first position measurement as close as possible to the interaction point, the active region begins 8 mm from the nominal centre of the beamline. This is narrower than the safe beam tolerance during injection, so the VELO is therefore split into two retractable halves, with the modules only moved in to their nominal position once the beams are stable. A small overlap between the two halves ensures full azimuthal coverage.

#### 3.2.2 The TT

A schematic diagram of the VELO is shown in Fig. 3.4. The TT subdetector consists of four layers of 183 µm pitch silicon microstrip detector, with strips in the two outer layers (x) oriented vertically, and the two inner layers (u, v) oriented at stereo angles of (-5°, +5°) to the vertical plane. The stereo angle allows the reconstruction of the hit position in the horizontal plane, giving a single hit spatial resolution of ~ 50 µm. In outer regions, where the occupancy is lower, multiple strips share the same readout channel, reducing the total number of channels.



Figure 3.5: Layout of the IT for an x layer (a) and a u layer (b). Figure taken from [29].



Figure 3.6: Layout of the OT (a) and end view of an individual OT module (b). Figure taken from [30].

#### 3.2.3 The IT

A schematic diagram of the VELO is shown in Fig. 3.5. Similarly to the TT, the IT consists of four layers of 193 µm pitched silicon microstrip detector. The strips are arranged in the same (x, u, v, x) pattern as the TT, with a ±5 ° stereo angle. As with the TT, the IT single hit spatial resolution is ~ 50 µm.

#### 3.2.4 The OT

A schematic diagram of the OT is shown in Fig. 3.6. The OT consists of 5 mm inner diameter straw drift tubes, arranged in the same (x, u, v, x) pattern as the TT and IT, with a  $\pm 5^{\circ}$  stereo angle. The geometry of an OT module is shown in Fig. 3.6. The gas used is a mixture of 70% Ar, 30% CO<sub>2</sub>, chosen to give a drift time of less than 50µs, giving a drift co-ordinate resolution of 200 µm.



Figure 3.7: Schematic diagram of the readout granularity of the ECAL (a) and HCAL (b). In both cases, the beam pipe is in the lower left corner. The SPD and PS share the same granularity as the ECAL. Figure taken from [31].

# **3.3** Calorimeters

The calorimeter is divided into two subdetectors: the Electromagnetic CALorimeter (ECAL) and the Hadronic CALorimeter (HCAL). The ECAL is situated immediately downstream of RICH2, the HCAL immediately downstream from the ECAL. Electromagnetic showers, initiated by electrons or photons, deposit the majority of their energy in the ECAL, whereas showers initiated by hadrons deposit the majority of their energy in the HCAL. Both are sampling calorimeters, formed from alternating layers of absorbing plates and scintillating tiles. The ECAL employs lead absorbing plates, with readout split into three regions of differing granularity, to account for differing occupancy with polar angle. Similarly, the HCAL employs iron absorbing plates, and is divided into two regions of granularity. The layout of both calorimeters is shown in Fig. 3.7. The ECAL is 25 radiation lengths thick, and has a Moliere radius of 2.5 cm. The HCAL is 5.9 nuclear interaction lengths thick, preceded by 1.1 nuclear interaction lengths in the ECAL.

Before the main body of the calorimeter, the ECAL incorporates a Scintillator Pad Detector (SPD) and Pre-Shower (PS). The SPD consists of a layer of scintillating pad, which identifies charged particles on entry to the ECAL and therefore allows electrons and photons to be distinguished using only calorimeter information. The PS consists of a 2.5 radiation length lead plate followed by a layer of scintillating pad, and allows a longitudinally segmented measurement of the energy deposits.

# **3.4** Charged hadron identification

The Ring Imaging CHerenkov (RICH) subdetectors measure Cherenkov radiation produced by charged particles, in order to distinguish between charged hadrons. Cherenkov radiation results from a particle travelling through a medium faster than the speed of light in that medium. This results in a cone of photons being emitted, with the angle



Figure 3.8: Polar angle versus momentum for pions from simulated  $B^0 \to \pi^+\pi^-$  events. The boxes indicate the angular acceptance and momentum coverage of the two RICH subdetectors. Figure taken from [32].



Figure 3.9: Schematic diagram of RICH1 (a) and RICH2 (b). Figure taken from [27].

of the cone dependent on the velocity of the particle. Measurement of this angle, along with the momentum measurements from the tracker, allows pions, kaons and protons to be distinguished. The requirement that the particle velocity exceeds the speed of light in the medium results in a threshold in momentum for a given medium. As the momentum increases, the angle of the cone increases, and it becomes more difficult to distinguish particles. A given material is therefore only useful for distinguishing between particles over a certain range of momenta.

To maximise the effectiveness of the RICH system, three different radiators are used: aerogel, which has a refractive index 1.03 (providing  $K - \pi$  separation in the 2 < P < 10 GeV/c range),  $C_4F_{10}$  gas, which has a refractive index 1.0014 (providing  $K - \pi$  separation in the 10 < P < 60 GeV/c range), and  $CF_4$  gas, which has a refractive index 1.0005 (providing  $K - \pi$  separation in the 15 < P < 100 GeV/c range). The RICH detector is split into two subdetectors: RICH1 and RICH2. The RICH1 subdetector contains the aerogel and  $C_4F_{10}$  radiators, and is situated between the VELO and TT. The RICH2 subdetector contains the  $CF_4$  radiator and is situated downstream of the T-stations, with a reduced geometric acceptance of 15 mrad to 100 (120) mrad in the vertical (horizontal) plane. This arrangement of subdetectors is driven by the distribution of high momentum particles, which are predominantly at low angles to the beamline, as shown in Fig. 3.8.

The structure of the two RICH subdetectors are shown in Fig. 3.9. The spherical mir-



Figure 3.10: Reconstructed Cherenkov angle in the  $C_4F_{10}$  radiator versus momentum for tracks with well isolated Cherenkov rings. Different hadron types form clearly distinguishable bands in certain momentum ranges. Figure taken from [33].

rors focus the cone of Cherenkov photons to form a ring on the photodetector planes. The reconstructed Cherenkov angle in the  $C_4 F_{10}$  radiator is shown against reconstructed track momentum in Fig. 3.10 for well isolated tracks taken from 2011 data [33]. Here, isolated tracks are defined as having a Cherenkov ring which does not overlap with any other Cherenkov rings in a given radiator. This isolation requirement ensures that few photons which are considered as being within the Cherenkov ring originate from other sources, giving a highly pure sample of Cherenkov photons. This high purity allows different hadron types to be clearly distinguishable in Fig. 3.10 for certain momentum ranges. In general, however, Cherenkov photons from a given track are not so easily distinguished from other, unrelated photons, and so a more sophisticated procedure is required to identify different hadron types. Using measurements of the track momentum and trajectory from the tracking system, the centre and radius of the Cherenkov ring formed on the photodetector plane can be calculated for a given mass hypothesis. The consistency of this hypothesis can then be tested by a fit of this expected ring to the measured Cherenkov photons. This fit is used to form a likelihood for each particle identity hypothesis. This information is used in analyses by taking the difference in log-likelihood (DLL) between two particle identity hypotheses. For example, for separating kaons from pions the DLL between the kaon and pion mass hypotheses is defined as:

$$DLL_{K\pi} = \log(\mathcal{L}_K) - \log(\mathcal{L}_{\pi}) \tag{3.1}$$



Figure 3.11: Schematic diagram of the muon system. The arrangement of the stations and absorber plates is shown in (a), with the four regions of readout granularity (b). Figures taken from Ref. [27](a) and from Ref. [34](b).

where  $\mathcal{L}_h$  represents the likelihood of the *h* mass hypothesis (increasing  $\mathcal{L}_h$  represents increasing consistency with the *h* hypothesis).

### 3.5 The muon system

Muons are highly penetrating particles, and traverse the calorimeters without depositing a significant fraction of their energy. This distinguishes muons from other charged particles, the vast majority of which deposit the entirety of their energy in the calorimeters. Measuring charged particles surviving beyond the calorimeter, and through additional shielding material, allows muons to be identified with a high purity. Due to their ease of identification and the presence of muons in many B decays of interest, the muon identification system is designed to be incorporated into the LHCb hardware trigger, as described in Sect. 3.10.

A schematic diagram of the muon system is shown in Fig. 3.11. The muon system consists of five stations, M1-M5. The first station, M1, is located between RICH2 and the ECAL, while M2-M5 are located downstream of the HCAL. Three 80 cm thick iron absorber plates, corresponding to a total of 20 nuclear interaction lengths, are interleaved with M2-M5, as shown in Fig. 3.11(a). The location of M1 allows a much greater baseline for the reconstruction of tracks using only the muon system, which is made use of in the hardware trigger (see Sect. 3.10). The muon system is instrumented with Multi-Wire

Track momentum	Hits required in muon stations
3 GeV/c < P < 6 GeV/c	M2 and M3
$6 \mathrm{GeV}/c < P < 10 \mathrm{GeV}/c$	M2, M3 and either M4 or M5
$P > 10 \mathrm{GeV}/c$	M2, M3, M4 and M5

Table 3.2: The isMuon identification requirements.

Track momentum	Hits required in muon stations
3 GeV/c < P < 6 GeV/c	at least two of M2, M3 and M4
P > 6  GeV/c	at least three of M2, M3, M4 and M5

Table 3.3: The isMuonLoose identification requirements.

Proportional Chambers (MWPCs) for most of the active area, and triple Gaseous Electron Multipliers (GEMs) in the inner region of M1 (region R1 in Fig. 3.11(b)). The MWPCs (GEMs) have a time resolution of 5ns (3ns), considerably smaller than the minimum bunch crossing spacing of 25ns.

Due to the strong suppression of other charged particles, a simple boolean identification criteria based on the presence of hits in the muon system provides good performance. Tracks are extrapolated into the muon system, and hits searched for within a field of interest opened around the track in each station. Muons with a low momentum may not traverse the entirety of the muon system, so the number of stations in which a hit is required varies as a function of momentum. The magnet prevents any muons with a momentum below 3 GeV/c from traversing the entire muon system. The main muon identification criteria, the isMuon requirement, is summarised in Tab. 3.2. A requirement giving a higher efficiency at the cost of reduced purity, isMuonLoose, is summarised in Tab. 3.3. In addition to these boolean requirements, further discrimination is provided by a DLL, based on the consistency of the hit positions with the extrapolated track. The consistency of the hits is quantified with a 'squared distance' variable, defined as:

$$D = \frac{1}{N} \sum_{i=0}^{N} \left\{ \left( \frac{x_{\text{closest},i} - x_{\text{track}}}{x_{\text{pad}}} \right)^2 + \left( \frac{y_{\text{closest},i} - y_{\text{track}}}{y_{\text{pad}}} \right)^2 \right\},\tag{3.2}$$

where  $x_{\text{closest,i}}$  and  $y_{\text{closest,i}}$  are the co-ordinates of the closest hit to the track in station i,  $x_{\text{track}}$  and  $y_{\text{track}}$  are the co-ordinates the extrapolated track, and  $x_{\text{pad}}$  and  $y_{\text{pad}}$  are the dimensions of the logical pad containing the hit. Only hits consistent with the isMuon-Loose requirement are included. This DLL from the muon system can be combined with the DLLs from the RICH, and calorimeter, giving a single combined DLL<sub> $\mu\pi$ </sub> containing all sources of PID information.



Figure 3.12: Schematic diagram showing the different track types. The magnetic field as a function of z position is shown above on the same scale, to give an indication of the integrated magentic field each track type traverses. Figure taken from [27].

# **3.6** Track reconstruction

Tracks may be formed from the information from different combinations of the tracking subdetectors, depending on the origin and trajectory of the particles. The following types of tracks are defined in the LHCb reconstruction.

- **VELO tracks** are formed solely from hits in the VELO modules. The particles traverse no significant magnetic field in the VELO, and the tracks are therefore straight and no momentum or charge measurement is possible.
- Upstream tracks are formed using hits from the VELO and TT subdetectors. For particles which are swept out of the T station acceptance by the magnetic field, upstream tracks are the longest track type reconstructable. While the tracks do

not cross the magnet, some field is present between the VELO and TT, allowing a momentum measurement with  $\sim 10\%$  precision.

- T-tracks are formed solely from hits in the T stations.
- Downstream tracks are formed from hits in the TT and T stations. For the decay products of particles which decay after the VELO (e.g.  $K_s^0$  decay products), downstream tracks are the longest track type reconstructable.
- Long tracks are formed from hits in the VELO and T stations. They are the default track type used in physics analyses, as they provide the best momentum resolution. Except where specifically stated otherwise, all tracks used in the remainder of this thesis are long tracks.

These track types are sketched in Fig. 3.12, with the magnetic field strength shown above the detector schematic.

# **3.7** Definition of reconstructed quantities

#### 3.7.1 Reconstructed track quantities

Tracks originating from beauty and charm hadron decays may be measured as being inconsistent with originating from the PV. This is determined by two quantities: the Impact Parameter (IP), defined as the distance of closest approach between the track and the PV, and the Impact Parameter  $\chi^2$  (IP $\chi^2$ ), defined as the change in the  $\chi^2$  of the PV fit when including the track. Typically the IP $\chi^2$  offers greater discriminating power between tracks originating from heavy flavour and from the PV, and so is more commonly used in analyses.

There are two types of misreconstructed tracks: ghost tracks, and clone tracks. Ghost tracks are formed from a collection of unrelated hits (possibly including track segments from one or more particles). They are rejected by the track  $\chi^2$  returned by the track hits. Clone tracks are formed by reconstructing multiple tracks originating from the same particle. They are identified using the Kullback-Leibler (KL) distance between track pairs, which quantifies the difference in information between the two tracks [35, 36]. Where the KL distance for a track pair is below 5000, only the track containing the greatest number of hits is kept.

#### 3.7.2 Reconstructed vertex quantities

Tracks originating from decays of short-lived particles should point back to a single origin point. This is quantified by fitting a vertex to the combination of tracks, and considering the  $\chi^2$  of the fit, denoted vertex  $\chi^2$ . The displacement between two vertices (e.g the vertex of a *B* meson decay and the PV) is quantified by the flight distance  $\chi^2$ , which is defined as the change in vertex  $\chi^2$  when both vertices are combined into a single vertex.

For fully reconstructed decays, the combined momentum vector of the decay products should point back to the PV. This is quantified as  $\theta_P$ , an angle between the reconstructed momentum and the displacement vector between the SV and PV. Additional information about the consistency of the decaying particle originating from the PV is provided by the  $IP\chi^2$  of the reconstructed momentum vector, which should be small for fully reconstructed decays, unlike in the case of the individual tracks.

# 3.8 Simulated events

Simulated events are used to model the selection efficiencies for signal decays, along with the kinematic distributions for many signal and background decays. To generate these, the full chain is simulated, from the production of particles in proton-proton collisions, to the decays of heavy flavoured hadrons, and the LHCb detector response to every particle in an event. Proton-proton collisions are simulated in PYTHIA [37], with a specific LHCb tuning [38]. Heavy flavoured particles are produced by PYTHIA, but their decays are modelled using the EVTGEN package [39], with final state radiation generated using PHOTOS [40]. The response of the LHCb detector to each particle is simulated in GEANT4 [41], as described in Ref. [42]. This complete simulation chain is encapsulated in a software package called GAUSS. The response of the LHCb subdetectors is then digitised using the BOOLE [43] software. After simulation, events are processed by the LHCb reconstruction in the exact same way as for data. The trigger algorithms are applied to the simulated data but the decisions are merely stored along with the data, rather than used to reject events.

The tuning of the LHCb simulation has changed over time, due to differences in running conditions and an improving understanding of the detector. The simulated events used in chapters 4, 5 and 6 correspond to the simulation conditions at the end of 2010, 2011 and 2012 respectively.

# 3.9 Performance of the LHCb detector

The efficiency to reconstruct tracks has been measured, using  $J/\psi \rightarrow \mu^+\mu^-$  decays selected from data and from simulated events. The ratio between the track reconstruction efficiency in data and simulated events is shown in Fig. 3.13 as a function of momentum and pseudorapidity. The data shown was taken in 2011. The average efficiency ratio between data and simulated events is 1.009  $\pm$  0.006. The momentum of tracks is measured with



Figure 3.13: The ratio of track reconstruction efficiency between 2011 data and simulated events, as a function of momentum and pseudorapidity. Figure taken from Ref. [44]



Figure 3.14: IP resolution in measured in data (black) and simulation (red), as a function of inverse  $p_{\rm T}$ . Figure taken from [45].

0.4% (0.6%) precision at momenta of 3 GeV/c (100 GeV/c). For tracks in the kinematic range covered by *B* decays, the charge mis-assignment rate is negligible.

The IP resolution in the vertical plane for tracks taken from 2012 data and from simulated events is shown in Fig. 3.14. The simulated events in Fig. 3.14, which are generated using the conditions from the end of 2012, show good agreement with the data. This is not the case for previous simulation conditions, such as those from 2011 and 2010. The poor agreement between the IP resolution in  $B^+ \rightarrow J/\psi K^+$  events in 2011 data and simulated events generated using 2011 conditions is shown in Fig. 3.15. For events generated under such conditions, a smearing tool was used to increase the IP resolution in simulated events to match data. Simulated events are shown in Fig. 3.15 before and after this smearing, demonstrating that the smearing tool produces good agreement.

Hadron PID efficiencies are measured in data, using  $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$  decays,



Figure 3.15: Comparisons between data and simulated  $B^+ \rightarrow J/\psi K^+$  events for  $B^+ \operatorname{IP} \chi^2$ . The unsmeared, shown in blue (medium grey), and smeared, shown in light grey, simulated events are compared to data, shown in black.



Figure 3.16: The PID efficiency for kaons, and the pion misidentification rate, as a function of momentum. Efficiencies are shown for requirements of  $\text{DLL}_{K\pi} > 0$  and  $\text{DLL}_{K\pi} > 5$ , using  $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$  decays selected from data taken in 2012. Figure taken from [46].



Figure 3.17: Efficiency of the isMuon requirement for muons (a), protons (b), pions (c), and kaons (d), as a function of momentum. Several ranges of  $p_{\rm T}$  are shown. Figure taken from Ref. [47].

which allow pions and kaons to be identified unambiguously without the use of PID. Similarly, the proton PID efficiencies are measured using  $\Lambda^0 \to p^+\pi^-$  decays. The efficiency to identify kaons, and corresponding pion misidentification rate, is shown in Fig. 3.16 as a function of hadron momentum. In addition to momentum, the PID efficiencies also depend on the pseudorapidity of the track, and on the total occupancy of the LHCb detector. For typical kaons from a *B* decay, the efficiency to identify a kaon is ~ 95%, for a ~ 10% pion misidentification rate. The hadron PID efficiencies are not well reproduced in simulation, and so are taken entirely from the  $D^{*+} \to (D^0 \to K^-\pi^+)\pi^+$  and  $\Lambda^0 \to p^+\pi^-$  calibration samples.

Muon PID efficiencies are measured in data, using a tag-and-probe method with  $J/\psi \rightarrow \mu^+\mu^-$  decays. The misidentification rates for hadrons are measured using the  $D^{*+} \rightarrow (D^0 \rightarrow K\pi)\pi$  and  $\Lambda^0 \rightarrow p^+\pi^-$  calibration samples described previously. The efficiency for the isMuon requirement is shown in Fig. 3.17, along with the misidentification rates for pions, kaons and protons. Typical muon PID efficiencies are ~ 98% for a misiden-

tification rate of  $\sim 1\%$ .

# 3.10 Trigger

The LHC is designed to have a bunch crossing rate of 40MHz. Given an event size of  $\mathcal{O}(100 \text{ kB})$ , storing every event would mean writing  $\mathcal{O}(\text{TB/s})$  to offline storage, which is unfeasible. In order to avoid this, while still taking advantage of the high interaction rate, data are stored in temporary buffers and a fast decision taken on which events are interesting and should be stored.

The LHCb trigger consists of three stages. The first, Level 0 (L0), runs on custom electronics on-detector hardware. New data are generated every bunch crossing, so its operation must take a short and constant period of time, and be synchronised with the LHC clock. Given these time constraints only the most basic reconstruction is possible. Only calorimeter and muon chamber information is used and no track reconstruction is performed at this stage.

The remainder of the trigger is split into two levels, High Level Trigger (HLT) 1 and 2. The HLT runs as software on a dedicated CPU farm and its operation is asynchronous with the LHC clock. The first level, HLT1, uses the information obtained from partial event reconstruction to reduce the event rate from  $\sim$ 1MHz to  $\sim$ 50kHz, in a time window of 30ms. The second level, HLT2, applies a full event reconstruction and uses the resulting information to reduce the event rate to  $\sim$  3kHz.

#### 3.10.1 Hardware trigger

Due to the stringent timing requirements, the full detector information cannot be used in the L0 trigger, and so the trigger algorithms are divided into three categories, each using a different set of subdetectors. The L0Calorimeter triggers use only information from the calorimeters, the L0 muon triggers use only the muon system, and the L0PileUp triggers use only information from the VELO. The L0PileUp triggers are not made use of in this thesis, and so are not described below.

#### L0Muon triggers

The L0Muon trigger algorithms build tracks using hits in all five muons stations, M1-M5. Hits are required to form a straight line in the vertical plane, in which tracks are not bent by the magnetic field, pointing back towards the nominal proton-proton interaction region. The transverse momentum  $(p_T)$  of each track is measured with a momentum resolution of ~ 25% by assuming it has originated from the nominal proton-proton interaction region. The search window for hits in the horizontal plane prevents tracks with a  $p_T$  of less than 500 MeV/c from being reconstructed, reducing the processing time. There are two selection



Figure 3.18: Efficiency for L0Muon and L0DiMuon as a function of  $J/\psi p_{\rm T}$  for  $B^+ \to J/\psi K^+$ . Taken from Ref. [48].

algorithms: L0Muon, which places a requirement on the  $p_{\rm T}$  of a single muon track, and L0DiMuon, which places a requirement on the product of the  $p_{\rm T}$ s of two muon tracks. The efficiency of the L0Muon and L0DiMuon algorithms has been measured for  $B^+ \rightarrow J/\psi K^+$ decays in data, with the results shown in Fig. 3.18. Typical efficiencies for B decays to dimuon final states are ~ 90%.

#### L0Calorimeter triggers

The L0Calorimeter algorithms select charged hadrons, electrons and photons based on their energy deposits in the calorimeters. The energy deposited in a two-by-two calorimeter cell zone is summed, and the transverse component calculated by assuming the particle originated from the nominal p - p interaction region. For electrons and photons only the energy deposited in the ECAL is considered, with electrons identified by the presence of a hit in the SPD region matching the calorimeter deposit. For charged hadrons, the energy deposited in the ECAL and HCAL is summed. There are three selection algorithms, L0Hadron, L0Electron and L0Photon, each with a different transverse energy requirement, and with the electron identification described above discriminating between L0Electron and L0Photon. The efficiency of the L0Hadron algorithm has been measured in several typical hadronic B and D decays in data, with the results shown in Fig. 3.19. Typical efficiencies are ~ 25% for hadronic B decays. The efficiency for a B decay to pass the L0 trigger due to the other B hadron in the event is also ~ 25%, doubling the efficiency for hadronic B decays to pass L0.



Figure 3.19: Efficiency for L0Hadron as a function of  $B p_{\rm T}$  for  $B^- \to D^0 \pi^-$  and  $B^0 \to D^- \pi^+$ , and as a function of  $D p_{\rm T}$  for  $D^0 \to K^- \pi^+$  and  $D^+ \to K^- \pi^+ \pi^+$ . Taken from Ref. [48].

#### 3.10.2 The first stage of the High Level Trigger

The CPU power available to the HLT is not sufficient to allow the tracking algorithms to be run at the full ~1MHz L0 output rate. To allow the use of full tracking information later in the trigger, HLT1 must therefore reduce the rate to a low enough level to allow the tracking algorithms to be run, by performing a more restricted track reconstruction. The most time consuming stage of the track reconstruction procedure is constructing long tracks; performing the VELO only tracking requires considerably less time. The HLT1 reconstruction sequence therefore begins by performing the VELO track reconstruction. Selected VELO tracks are extrapolated into the T-stations, and hits searched for within given search windows. Requirements on the p and  $p_{\rm T}$  of the tracks limit the search windows, reducing the processing time. At this stage, the number of long tracks created is sufficiently low to allow the tracks to be fitted using a kalman filter based technique [49,50]. A restriction is then placed on the track  $\chi^2$ .

The analyses presented in this thesis use two HLT1 algorithms: HltTrackAllL0, and Hlt1TrackMuon. The two algorithms differ in the requirements placed on VELO tracks, and in the p,  $p_{\rm T}$  and track  $\chi^2$  requirements placed on long tracks. In both cases, VELO tracks are required to have a minimum of 9 hits present, a maximum of 2 hits along the interpolated track trajectory missing, and an IP greater than 0.1mm. In the Hlt1TrackMuon case, muon ID is required, which reduces the number of tracks sufficiently to allow looser p, and  $p_{\rm T}$  requirements than for HltTrackAllL0. The efficiency of the Hlt1Track algorithms



Figure 3.20: Efficiency for Hlt1TrackMuon as a function of  $J/\psi p_T$  for  $B^+ \to J/\psi K^+$ . Taken from Ref. [48].



Figure 3.21: Efficiency for Hlt1TrackAllL0 as a function of  $B p_{\rm T}$  for  $B^- \to D^0 \pi^-$  and  $B^0 \to D^- \pi^+$ , and as a function of  $D p_{\rm T}$  for  $D^0 \to K^- \pi^+$  and  $D^+ \to K^- \pi^+ \pi^+$ . Taken from Ref. [48].

has been measured for  $B^+ \to J/\psi K^+$ , and for several typical hadronic B and D decays in data, with the results shown in Fig. 3.20 and 3.21. Typical efficiencies are ~ 90% for B decays to dimuon final states, ~ 80% for hadronic final states.

#### 3.10.3 The second stage of the high level trigger

The HLT1 output rate of  $\sim$ 50kHz is sufficiently low to allow full track reconstruction. Due to the availability of full reconstruction, and the flexibility of a trigger implemented in software, HLT2 features a large number of algorithms. Algorithms are divided between those which select inclusively, based on a subset of the final state particles, and those which select a single final state exclusively.

The most general inclusive trigger is the 'topological trigger' which selects B hadron decays with at least two charged decay products. The topological trigger begins by forming two track combinations ('topo candidates'), with one track required to have satisfied Hlt2TrackAllL0 or Hlt1TrackMuon. A requirement is placed on a "corrected mass" variable originating from the SLD experiment [51]:

$$M_{corr} = \sqrt{m^2 + |p'_{\rm T}|^2} + |p'_{\rm T}|^2, \qquad (3.3)$$

where *m* is the reconstructed mass of the candidate and  $|p'_{\rm T}|$  is the momentum transverse to the direction of *B* meson travel, which is determined from the measured PV and *B* 



Figure 3.22: Efficiency for one of the Hlt2TopoBBDTnBody or Hlt2TopoMuBBDTnBody (n=2,3,4) algorithms as a function of  $B p_{\rm T}$  for  $B^+ \rightarrow J/\psi K^+$ . The efficiency is measured relative to events which have passed Hlt1TrackAllL0 or Hlt1TrackMuon. Taken from Ref. [48].



Figure 3.23: Efficiency for one of the Hlt2TopoBBDTnBody (n=2,3,4) algorithms as a function of  $B p_{\rm T}$  for  $B^- \rightarrow D^0 \pi^-$  and  $B^0 \rightarrow D^- \pi^+$ . The efficiency is measured relative to events which have passed Hlt1TrackAllL0. Taken from Ref. [48].

candidate decay positions, and from the measured B candidate momentum. This corrected mass variable corresponds to the mass of the decaying hadron given a missing massless particle with zero momentum along the axis of B travel. Tracks are required to have an  $IP\chi^2$  greater than 4, and a distance of closet approach (DOCA) of less than 0.2 mm. If any additional tracks can be added using the same criteria, three or four track topo candidates are also built. These topo candidates are then required to pass a requirement on the output of a multivariate classifier, which is designed to be robust against variations in decay topology or in detector calibration [52]. Inputs to the classifier include kinematic properties such as the scalar sum of the track  $p_{\rm T}$  and the (corrected) mass of the n track combination, and geometric properties such as the flight distance  $\chi^2$  of the candidate, and the IP $\chi^2$ s and DOCAs of the tracks. This defines the Hlt2TopoBBDTnBody trigger algorithms, for n = (2,3,4) body combinations. In addition to these algorithms, if one of the tracks passes the isMuon requirement a loosened requirement on the MVA classifier is applied, defining the Hlt2TopoMuBBDTnBody algorithms. The efficiency of the topological trigger algorithms has been measured for  $B^+ \to J/\psi K^+$ , and for several typical hadronic B decays in data, with the results shown in Fig. 3.22 and 3.23. Typical HLT2 efficiencies are  $\sim 85\%$ for B decays to dimuon final states,  $\sim 70\%$  for hadronic final states.

Hadronic decays of charmed mesons are selected using exclusive triggers. The decay  $D^0 \to K^-\pi^+$  is selected using the Hlt2CharmHadD02HH\_D02KPi algorithm. This algorithm requires both tracks to have a track  $\chi^2$  below three, and an IP $\chi^2$  greater than 9. In addition, tracks are required to have a momentum greater than 5 GeV/*c*, and a transverse momentum greater than 1500 MeV/*c*. The  $D^0$  candidate is required to have a vertex  $\chi^2$  below 10, a flight distance  $\chi^2$  above 40, and a transverse momentum greater than 2000 MeV/*c*. The  $D^0$  candidate is also required to have an invariant mass within 50 MeV/ $c^2$  of the nominal  $D^0$  mass, and to point back to the primary vertex with a requirement of  $\cos(\theta_P)$  above 0.99985. This pointing requirement is sufficiently loose to allow  $D^0 \to K^-\pi^+$  decays originating from *B* meson decays to be selected by this trigger algorithm. This fact is exploited in Chapter 6 to select the  $B^0 \to D^{*-}\ell^+\nu$  decay, which suffers from a kinematic bias when selected using other trigger algorithms.

For the entire trigger system, typical efficiencies are  $\sim 80\%$  for *B* decays to dimuon final states,  $\sim 40\%$  for hadronic final states.

## 3.11 Multivariate selection

Many different variables offer discriminating power between signal and backgrounds, and so any selection technique should be able to make use of multiple sources of information. The simplest method to select events based on multiple variables is simply to impose a requirement on each variable individually, a procedure known as cut-based analysis. This does not use the full information however, as the correlations between variables are ignored. Where these correlations are different between signal and background events, they may provide considerable discriminating power. In order to make use of this information a multivariate analysis (MVA) is used. An MVA classifier is a function transforming multiple input variables into a smaller number of output variables (MVA output). In all cases in this thesis, a single output variable is produced. The function is designed such that the distributions in MVA output differ as greatly as possible between signal and background. Machine learning techniques are used to generate the MVA, using samples of events identified as signal and background (training samples). The MVA output can then be used to discriminate between signal and background events, based on their resemblance to the signal or background training samples. The training samples should resemble the actual signal and background events as closely as possible, or the separation between signal and background in MVA output will be suboptimal.

Many different MVA algorithms exist, differing in both the functional form of the MVA classifier itself, and the machine learning technique used to generate it. In situations where the input variables have simple mathematical relationships, the relative performance of these algorithms have been extensively studied. In cases where the relationships between variables are less simple, however, the situation is less clear. Empirically, a Boosted Decision Tree (BDT) [53] is seen to give the best performance in all cases studied in this thesis. A BDT takes a set of relatively weak individual classifiers (decision trees), and combines them into a single function offering far greater power.

A decision tree follows the structure shown in Fig. 3.24. At each node, the sample is split in a single variable. The variable and splitting point for each node is chosen to maximise the separation between signal and background, as determined by a figure of merit. The process continues until no sufficiently strong splitting can be made, the remaining sample contains too few events, or a predetermined maximum depth is reached. The lowermost nodes in the tree ('leaf nodes') are each classified as being signal-like or background like, based on the population of training events found at that node.

The decision tree generation process is then repeated, with the training events misclassified by the previous iteration assigned a higher weight in the next iteration (boosting). Various algorithms exist for assigning the weights, and for combining the output of the multiple decision trees generated into a single function. In this thesis, both the adaptive boosting (AdaBoost) [54] and gradient boosting (GradBoost) algorithms are used. The choice between the two is made purely empirically for each application. In all cases, the TMVA package [55] is used to generate the MVA.



Figure 3.24: Schematic view of a decision tree.

# Chapter 4

# Search for the lepton number violating decay $B^+ \rightarrow h^- \mu^+ \mu^+$

# 4.1 Introduction

In the SM neutrinos are massless, with only left-handed neutrinos and right-handed antineutrinos existing. The presence of neutrino oscillations requires the addition of neutrino mass terms to the SM Lagrangian. There are two types of mass terms which can be added, Dirac and Majorana masses [56]. Models containing both types of mass terms require the introduction of one or more right-handed gauge singlet fermion fields [57]. These carry no charge for any of the SM forces and their only interaction is via mixing with the SM neutrinos. They are therefore often referred to as sterile neutrinos. The number and masses of these particles is, in general, poorly constrained, and a large variety of models exist [58].

The existence of Majorana neutrinos would imply that lepton number, L, is no longer conserved and it would be possible to have decays which violate lepton number by two units ( $\Delta L = 2$ ) [58]. The most well known experimental signature for such processes is neutrinoless double beta decay [59]. However, as seen in Fig. 4.1,  $\Delta L = 2$  processes could also be observed through meson decays of the form  $X^+ \to Y^- \ell^+ \ell^+$ . The rate of such processes is highly suppressed by the small sterile-SM neutrino mixing but can be substantially enhanced if the neutrino is on the mass shell [4].

In this chapter a search for the  $\Delta L = 2$  decay  $B^+ \to h^- \mu^+ \mu^+$  is presented, where  $h^$ represents a  $\pi^-$  or a  $K^-$ . The decay is sensitive to sterile neutrinos with masses in the range  $(m_h + m_\mu) < m < (m_B - m_\mu)$ . While the range of neutrino masses accessible via Bmeson decays is relatively narrow, this range is favoured in the  $\nu MSM$  model introduced in Sect. 2.3.2. In order to probe theoretically allowed mixing angles,  $B^+ \to K^- \mu^+ \mu^+$ and  $B^+ \to \pi^- \mu^+ \mu^+$  must be measured beyond branching fractions of  $3.6 \times 10^{-14}$  or  $6.3 \times 10^{-13}$  respectively [4]. However, in order to probe theoretically allowed mixing angles,  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$  must be measured beyond branching fractions of



Figure 4.1: Generic s-channel Feynman diagram for  $\Delta L = 2$  meson decays. The s-channel makes the dominant contribution in processes with a sterile neutrino on the mass-shell [4].

 $3.6 \times 10^{-14}$  or  $6.3 \times 10^{-13}$  respectively [4]. Prior to the measurement presented in this chapter, the best experimental limit on the  $B^+ \to K^- \mu^+ \mu^+$  ( $B^+ \to \pi^- \mu^+ \mu^+$ ) branching fraction was  $\mathcal{B}(B^+ \to K^- \mu^+ \mu^+) < 1.8 \times 10^{-6}$  ( $\mathcal{B}(B^+ \to \pi^- \mu^+ \mu^+) < 1.2 \times 10^{-6}$ ) at 90% confidence level [60].

The search is performed using pp collision data, corresponding to an integrated luminosity of ~ 36 pb<sup>-1</sup>, collected with the LHCb detector during 2010. The criteria described in Sect. 4.2 are used to select  $B^+ \to h^- \mu^+ \mu^+$  candidates while suppressing both combinatorial background and genuine B decays with final state particles misidentified. The number of combinatorial background events in the signal window is estimated from the  $M_{h^-\mu^+\mu^+}$  sideband. The number of misidentified B decay candidates in the signal window is estimated using samples of simulated decays, scaled by branching fractions taken from Ref. [61] and PID efficiencies measured from data. The misidentified backgrounds, along with the measurements of PID efficiencies, are described in Sect. 4.3. The  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$  branching fractions are measured relative to that of  $B^+ \to J/\psi (\to \mu^+ \mu^-)K^+$  (hereafter denoted  $B^+ \to J/\psi K^+$ ), as described in Sect. 4.4. The limit setting procedure is described in Sect. 4.5, and the treatment of systematic uncertainties in Sect. 4.6. Finally, the results are presented in Sect. 4.7. This measurement has been published in Ref. [1].

# 4.2 Signal selection and combinatoric background

The  $B^+ \to \pi^+ \mu^+ \mu^-$ ,  $B^+ \to K^+ \mu^+ \mu^-$ ,  $B^+ \to \pi^- \mu^+ \mu^+$  and  $B^+ \to K^- \mu^+ \mu^+$  candidates are selected by combining pairs of muons with a charged pion or kaon. The individual particles and the resulting *B* candidate are required to pass the set of preselection criteria that are given in Tab. 4.1. These preselection criteria serve to reduce the size of the dataset to a level the LHCb computing framework can accommodate.

The selection is optimised entirely on  $K^+\mu^+\mu^-$  candidates in data, leaving the  $h^+\mu^+\mu^+$ 

Cut	Value
$\mu^+$ track $\chi^2/ndof$	< 5
$\mu^+ p$	$> 3 \mathrm{GeV}/c^2$
$\mu^+ p_T$	$> 500 \mathrm{MeV}/c^2$
$\mu^+ \chi^2_{IP}$	> 4
$h^+$ track $\chi^2/ndof$	< 5
$h^+ p$	$> 2 \mathrm{GeV}/c$
$h^+ p_T$	$> 300 \mathrm{MeV}/c$
$K^+$ DLL $_{K\pi}$	> -1
$h^+ \chi^2_{IP}$	> 4
$B^+ \theta_P$	$< 63 \mathrm{mrad}$
$B^+$ vertex $\chi^2/\text{ndof}$	< 10
$m_B$	$  5.050 < m_{K\mu\mu} < 5.780 \text{GeV}/c^2  $

Table 4.1: The selection criteria used in the preselection.

candidates completely unbiased. The opposite-sign  $(K^-\mu^+\mu^-)$  final state, where the muons are required to have  $|m_{\mu^+\mu^-} - m_{J/\psi}| < 50 \text{ MeV}/c^2$ , is used as a proxy for the signal decay. The total number of  $B^+ \to J/\psi K^+$  candidates remaining after the application of the optimised selection requirements is  $\mathcal{O}(1000)$ . This number is large enough such that the yield of  $B^+ \to J/\psi K^+$  is still sufficiently unbiased to be used to normalise the  $B^+ \to h^-\mu^+\mu^+$  branching fraction, as described in Sect. 4.4.

The opposite-sign final state, where the muons are required to have  $m_{\mu^+\mu^-}$  outside of the ranges  $3600 < m_{\mu^+\mu^-} < 3740 \text{ MeV}/c^2$  and  $2900 < m_{\mu^+\mu^-} < 3200 \text{ MeV}/c^2$ , is used as a sample of background events.

The selection requirements were optimised to maximise  $S/\sqrt{B}$ , where S is the number of signal events and B is the number of combinatorial background events expected in the signal window. Two alternative figures of merit, S/B and  $S/\sqrt{(S+B)}$  were also considered, and did not result in a different choice of selection. The optimisation procedure was limited by the low number of background events remaining after the application of the selection criteria. The selection criteria chosen are listed in Tab. 4.2. The  $K^+\mu^+\mu^-$  final state does not allow the PID requirement used to select  $B^+ \to \pi^-\mu^+\mu^+$  to be optimised. The pion identification requirement is therefore chosen to be  $DLL_{K\pi} < -1$ .

To reject background coming from two neighbouring tracks that are close together in the tracking system and can share hits in the muon system, a cut is made on the number of shared hits in the muon system, nShared. In order to avoid selecting a muon as the pion or kaon, the hadron is required to be within the acceptance of the muon system (inMuonAcceptance flag=1), and to have insufficient muon system hits to be consistent with a muon (isMuonLoose flag=0).

Selection	Criteria		
Kinematic Selection:			
$m_B$	$5.050 < m_{K\mu\mu} < 5.780 \text{GeV}/c^2$		
track $p_{\rm T}$	$> 0.8  { m GeV}/c$		
track $\chi^2_{IP}$	> 45		
$\mu^+$ nShared	< 3		
$B^+ \chi^2_{IP}$	< 9		
$B^+ p_{\rm T}$	$> 2.5 \mathrm{GeV}/c$		
$B^+$ vertex $\chi^2/\text{ndof}$	< 4		
$B^+ \theta_P$	$< 8 \mathrm{mrad}$		
$B^+$ flight distance $\chi^2$	> 144		
$h^+ p$	$> 4 \mathrm{GeV}/c$		
PID Requirements:			
$K^+(\pi^+)$ DLL <sub><math>K\pi</math></sub>	> 1(< -1)		
$\mu^+ \text{ DLL}_{\mu\pi}$	> 3		
$\mu^+$ is Muon	True		
$h^+$ is MuonLoose	False		
$h^+$ in MuonAcceptance	True		

Table 4.2: The selection criteria.



Figure 4.2: The  $M_{K^+\mu^+\mu^-}$  mass distribution of  $K^+\mu^+\mu^-$  events after the application of the selection criteria (a) selecting the  $J/\psi$  invariant mass window, and (b) excluding the  $J/\psi$  and  $\psi(2S)$  invariant mass windows. The fit model is described in the text.

#### 4.2.1 Control channels

In Fig. 4.2(a), the  $m_{K^+\mu^+\mu^-}$  invariant mass distribution for events with  $|m_{\mu^+\mu^-} - m_{J/\psi}| < 50 \text{ MeV}/c^2$  is shown, after the application of all selection criteria. The  $B^+ \to J/\psi K^+$  mass peak is fitted with a Crystal Ball [62] Probability Density function (PDF) to account for the small radiative tail. The combinatorial background is fitted as flat line, and the partially reconstructed events in the lower mass sideband are fitted with a Gaussian PDF. All fits are performed using an unbinned maximum likelihood method. The  $B^+ \to J/\psi K^+$  peak in data has a Gaussian component of width 20 MeV/c<sup>2</sup> and the signal window is therefore defined as  $5280 \pm 40 \text{ MeV}/c^2$ . The  $B^+ \to J/\psi K^+$  peak contains  $3642 \pm 61$  signal events in total,  $3407 \pm 57$  within the above signal window. After tuning the selection criteria, the entire upper-mass sideband contained three events.

The invariant distribution for with  $m_{K^+\mu^+\mu^-}$ mass events  $|m_{\mu^+\mu^-} - m_{J/\psi,\Psi(2S)}| > 70 \,\text{MeV}/c^2$  is shown in Fig. 4.2(b). The events in the  $5280 \pm 40 \,\mathrm{MeV}/c^2$  mass window were not used to optimise the selection and are completely unbiased. Using the same fit model as previously described, with all shape parameters fixed to those from the above fit, the signal peak was determined to contain  $27 \pm 5$  events from the  $B^+ \to K^+ \mu^+ \mu^-$  decay. The observation of this decay indicates that the selection criteria do not exclusively select events with a dimuon mass close to the  $J/\psi$ , and that the efficiency is high enough to observe a decay with a branching fraction of  $\mathcal{O}(10^{-7})$ .

#### 4.2.2 Combinatorial background estimation

The expected number of combinatorial background candidates in the signal window,  $n_{comb}$ , is determined by extrapolating from the number of candidates observed in the upper mass sideband,  $n_{side}$ , using a scale factor  $n_{comb} = n_{side}/\tau$ . Due to the low number of events in the upper mass sidebands, no fit is performed to the invariant mass distribution. A range of scale factors,  $2 < \tau < 4.75$ , are considered for the limit setting procedure described in Sect. 4.5. The lower bound on  $\tau$  assumes the combinatorial background is flat with  $M_{h\mu\mu}$ . The upper limit is chosen to be greater than the highest slope seen in the  $K\mu\mu$  final state, using loosened selection requirements.

# 4.3 Peaking backgrounds

Fully reconstructed *B* hadron decays with one or more particles misidentified can form a background to the  $B^+ \rightarrow h^- \mu^+ \mu^+$  final state. Expectations for these backgrounds are computed using simulated events, and branching fractions taken from Ref. [61]. These backgrounds fall into two classes: decays with two hadrons misidentified as muons (e.g  $B^+ \to \pi^+ \pi^+ \pi^-$ , with two pions misidentified as muons); and decays with a muon misidentified as a hadron, and a hadron misidentified as a muon (e.g  $B^+ \to J/\psi K^+$ , with the kaon misidentified as a muon and a muon misidentified as a hadron). Computing the expectations for the first class requires the efficiencies for  $K^+ \to \mu^+$  and  $\pi^+ \to \mu^+$ , which are determined in Sect. 4.3.1. In addition to the above efficiencies, the second class also requires the efficiencies for  $\mu^+ \to K^+$  and  $\mu^+ \to \pi^+$ , which are determined in Sect. 4.3.1. Decays containing a misidentified proton are found to be negligible.

#### 4.3.1 Mis-id probabilities

#### $K \leftrightarrow \pi, \pi \rightarrow \mu$ and $K \rightarrow \mu$ mis-id probabilities

The  $K^+ \to K^+$ ,  $K^+ \to \mu^+$ ,  $\pi^+ \to \pi^+$ , and  $\pi^+ \to \mu^+$  efficiences are taken from a sample of  $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$  decays selected from the data. This sample allows kaons and pions to be identified unambiguously, without the use of PID. Two constraints are used to determine the correct mass assignments for the three charged particles in the  $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$  decay: the mass of the  $D^0$  candidate, and the mass difference between the  $D^{*+}$  and  $D^0$  candidates. Where the masses are correctly assigned the distributions of these quantities will form a narrow peak around the measured  $D^0$  and  $D^{*+}$  masses, whereas for candidates are required to have a value of  $M_{K^-\pi^+\pi^+} - M_{K^-\pi^+}$  close to the nominal  $D^0$  and  $D^{*+}$  mass difference. A fit to the  $M_{K^-\pi^+}$  distribution is then used to determine the number of candidates with correct mass assignments.

The trigger has a greater efficiency for particles identified as muons, due to trigger algorithms which require muon identification. For computing the  $\pi^+ \to \mu^+$  and  $K^+ \to \mu^+$ efficiences, the hadron is therefore required to have passed the trigger independently of such algorithms. No  $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$  events are selected by algorithms requiring hadron PID.

The efficiences for a given PID requirement are determined from a fit to the  $M_{K^-\pi^+}$  distribution, before and after application of the requirement in question. The signal is fitted with a Gaussian PDF, the combinatorial background with a second order polynomial. This procedure is performed in bins of P,  $P_T$  and track multiplicity.

#### $\mu^+\!\rightarrow K^+$ and $\mu^+\!\rightarrow \pi^+$ mis-id probabilities

The final two mis-id probabilities,  $\mu^+ \to K^+$  and  $\mu^+ \to \pi^+$  were evaluated using the muon tag and probe sample. This sample contains only muon candidates within the acceptance of the muon system. No information can be extracted about the (potentially much larger) mis-id rate outside this region. The selection outlined in Sect. 4.2 therefore required that the kaon/pion, which could be a mis-identified muon, traversed the muon detector acceptance. In this sample one muon has PID requirements applied (the isMuon flag is required), the "tag". The other has no PID requirements and is the "probe". The selection requirements for a muon given in Tab. 4.2 are applied to the probe. As above, the probe is required to have passed the trigger independently of algorithms requiring muon identification. No additional selection requirements have been imposed on the probe, beyond those given in Sect. 4.2. The  $M_{\mu^+\mu^-}$  distribution is shown in Fig. 4.3(a) after the application of these requirements. The  $J/\psi$  mass peak has been fitted with a Gaussian signal model, the background with a second order polynomial.

Muons mis-identified as kaons must have both isMuonLoose to be false, and  $DLL_{K\pi} > 1$ , as listed in Tab. 4.2. The fit to the  $M_{\mu^+\mu^-}$  distribution is repeated with these conditions required for one of the two muons in the sample. The effect on the mass distribution of requiring isMuonLoose to be false is shown in Fig. 4.3(b). The effect of requiring  $DLL_{K\pi} > 1$  is shown in Fig. 4.3(c). The effect of imposing both requirements together is shown in Fig. 4.3(d). The ratio between the yields before and after the application of both criteria gives the  $\mu^+ \rightarrow K^+$  mis-id rate,  $0.00 \pm 0.14\%$ . The same approach is used to evaluate the  $\mu^+ \rightarrow \pi^+$  mis-id rate, which was determined to be  $0.22 \pm 0.16\%$ . The error on these mis-id fractions dominates the uncertainty on the peaking background expected in the signal region. Due to the small number of candidates in the sample, no binning in kinematic variables is possible. However, the high (> 50 gevc) and low (< 10 GeV/c) momentum ranges were examined, in case the mis-id is much more prominent at either extreme. In both cases, the fitted signal with both mis-id requirements was consistent with zero.

If the two selection criteria used to reduce the  $\mu^+ \to K^+$  mis-id rate are assumed to be independent then the mis-id rate can be cross-checked by taking the product of the DLL<sub>K $\pi$ </sub> > 1 and isMuonLoose efficiency. This results in an estimate of  $0.02 \pm 0.03\%$ , which is consistent with the above result.

#### 4.3.2 Total misidentified background

The misidentifed backgrounds to  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$  are summarised in Tab. 4.3 and Tab. 4.4, respectively. The branching fractions for each mode are given, all of which are taken from Ref. [61]. The numbers of candidates expected in the total mass region and the signal window after the application of all selection requirements are then given. The mass distributions of these backgrounds are shown in Fig. 4.4. The dominant uncertainties are the branching fractions for each decay, and the  $\mu^+ \to K^+$  and  $\mu^+ \to \pi^+$  identification efficiencies. The total misidentifed background expected in the  $B^+ \to K^- \mu^+ \mu^+$  signal region is  $3.4^{+14}_{-0.2} \times 10^{-3}$  candidates. In the  $B^+ \to \pi^- \mu^+ \mu^+$  signal region,  $(2.9 \pm 0.6) \times 10^{-2}$  candidates are expected.



Figure 4.3: The  $M_{\mu^+\mu^-}$  distribution of tag and probe events after the application of the kinematic selection and trigger requirements: (a) with no additional requirements, (b) requiring  $\text{DLL}_{K\pi} > 1$ , (c) requiring isMuonLoose = 0, and (d) requiring both  $\text{DLL}_{K\pi} > 1$  and isMuonLoose = 0. The fit model is described in the text.



Figure 4.4: The  $M_{h^-\mu^+\mu^+}$  distributions for all of the misidentifed backgrounds considered, for  $B^+ \to K^-\mu^+\mu^+$  (left) and  $B^+ \to \pi^-\mu^+\mu^+$  (right), scaled to their respective expectations. The vertical lines indicate the signal window.

Mode	$\mathcal{B}$	Total	Signal window
		(candidates)	(candidates)
$B^+ \to D^0 \{ \to K^+ \pi^- \} \pi^+$	$(2.0 \pm 0.1) \times 10^{-4}$	$(4.3 \pm 0.2) \times 10^{-3}$	$(2.7 \pm 0.1) \times 10^{-4}$
$B^+ \to K^+ K^- K^+ [\mu^+ K^- \mu^+]$	$(3.4 \pm 0.2) \times 10^{-5}$	$(3.5 \pm 0.2) \times 10^{-2}$	$(1.3 \pm 0.1) \times 10^{-5}$
$B^+ \to \pi^+ K^- K^+ \ [\mu^+ K^- \mu^+]$	$(5.0 \pm 0.7) \times 10^{-6}$	$(3.6 \pm 0.5) \times 10^{-3}$	$(6.1 \pm 0.8) \times 10^{-4}$
$B^+ \to \pi^+ \pi^- K^+ \ [\mu^+ K^- \mu^+]$	$(5.1 \pm 0.3) \times 10^{-5}$	$(2.5 \pm 0.2) \times 10^{-3}$	$(1.7 \pm 0.1) \times 10^{-3}$
$B^+ \to \pi^+ \pi^- \pi^+ [\mu^+ K^- \mu^+]$	$(1.5 \pm 0.1) \times 10^{-5}$	$(1.7 \pm 0.2) \times 10^{-4}$	$(3.7 \pm 0.3) \times 10^{-4}$
$B^+ \rightarrow p\overline{p}K^+ \ [\mu^+K^-\mu^+]$	$(5.9 \pm 0.5) \times 10^{-6}$	$(1.0 \pm 0.3) \times 10^{-6}$	$0^{+8}_{-0} \times 10^{-8}$
$B^+ \rightarrow p\overline{p}\pi^+ \ [\mu^+ K^- \mu^+]$	$(1.6 \pm 0.2) \times 10^{-6}$	$(6.8 \pm 3.4) \times 10^{-6}$	$0^{+9}_{-0}  imes 10^{-7}$
$B^+ \to J/\psi \{ \to \mu^+ \mu^- \} K^+$	$(6.1 \pm 0.2) \times 10^{-5}$	$0.0^{+0.016}_{-0}$	$0.0^{+0.02}_{-0}$
$[\mu^+ K^- \mu^+]$			
$B^{0} \to J/\psi \{ \to \mu^{+}\mu^{-} \} K^{*0} \{ \to K^{+}\pi^{-} \}$	$(8.0 \pm 0.4) \times 10^{-5}$	$0.0^{+4.7}_{-0} \times 10^{-3}$	$0.0^{+3.2}_{-0} \times 10^{-4}$
$B_s^0 \to J/\psi \{ \to \mu^+ \mu^- \} \phi \{ \to K^+ K^- \}$	$(7.8 \pm 2.4) \times 10^{-5}$	$(3.5\pm0.8)\times10^{-3}$	$(4.5 \pm 1.9) \times 10^{-4}$
All	$5.3 \times 10^{-4}$	$4.9^{+1.6}_{-0.3} \times 10^{-2}$	$3.4^{+1.4}_{-0.2} \times 10^{-3}$

Table 4.3: A summary of misidentifed backgrounds for  $B^+ \to K^- \mu^+ \mu^+$ . The PID assignments required for each decay mode to be identified as  $B^+ \to K^- \mu^+ \mu^+$  are given in square brackets, with the order of the final state particles the same as in the original decay.

# 4.4 Normalisation

The  $B^+ \to h^- \mu^+ \mu^+$  branching fractions are measured relative to that of  $B^+ \to J/\psi K^+$ . The  $B^+ \to h^- \mu^+ \mu^+$  branching fraction is given by

$$\mathcal{B}(B^+ \to h^- \mu^+ \mu^+) = \frac{\mathcal{B}(B^+ \to J/\psi K^+)}{N_{B^+ \to J/\psi K^+}} \cdot \frac{\epsilon_{B^+ \to J/\psi K^+}}{\epsilon_{B^+ \to h^- \mu^+ \mu^+}} \cdot N_{B^+ \to h^- \mu^+ \mu^+}, \qquad (4.1)$$

$$= \alpha \cdot N_{B^+ \to h^- \mu^+ \mu^+}, \qquad (4.2)$$

where  $\mathcal{B}(X)$ ,  $N_X$  and  $\epsilon_X$  are the branching fraction, the number of events and the total efficiency, respectively, for decay mode X, and  $\alpha$  is the single event sensitivity.

The total efficiency to select  $B^+ \to h^- \mu^+ \mu^+$  events depends on the kinematics of the decay, which will depend on the physics which mediates the decay. The results are computed under two different assumptions.

Mode	B	Total	Signal window
		(candidates)	(candidates)
$B^+ \to D^0 \{ \to K^+ \pi^- \} \pi^+$	$(2.0 \pm 0.1) \times 10^{-4}$	$(4.9 \pm 0.2) \times 10^{-2}$	$(5.1 \pm 0.3) \times 10^{-3}$
$B^+ \to K^+ K^- K^+ \ [\mu^+ \pi^- \mu^+]$	$(3.4 \pm 0.2) \times 10^{-5}$	$(6.1 \pm 0.4) \times 10^{-4}$	$(8.1 \pm 0.5) \times 10^{-8}$
$B^+ \to \pi^+ K^- K^+ \ [\mu^+ \pi^- \mu^+]$	$(5.0 \pm 0.7) \times 10^{-6}$	$(2.1 \pm 0.3) \times 10^{-4}$	$(1.2 \pm 0.2) \times 10^{-7}$
$B^+ \to \pi^+ \pi^- K^+ \ [\mu^+ \pi^- \mu^+]$	$(5.1 \pm 0.3) \times 10^{-5}$	$(3.8 \pm 0.2) \times 10^{-2}$	$(6.5 \pm 0.4) \times 10^{-3}$
$B^+ \to \pi^+ \pi^- \pi^+ \ [\mu^+ \pi^- \mu^+]$	$(1.5 \pm 0.1) \times 10^{-5}$	$(1.5 \pm 0.1) \times 10^{-4}$	$(1.0 \pm 0.1) \times 10^{-2}$
$B^+ \rightarrow p\overline{p}K^+ \ [\mu^+\pi^-\mu^+]$	$(5.9 \pm 0.5) \times 10^{-6}$	$(3.3 \pm 1.0) \times 10^{-7}$	$0.0^{+8}_{-0} \times 10^{-8}$
$B^+ \rightarrow p\overline{p}\pi^+ \ [\mu^+\pi^-\mu^+]$	$(1.6 \pm 0.2) \times 10^{-6}$	$(1.4 \pm 0.8) \times 10^{-6}$	$(1.6 \pm 1.6 \times 10^{-7})$
$B^+ \to J/\psi \{ \to \mu^+ \mu^- \} K^+$	$(6.1 \pm 0.2) \times 10^{-5}$	$(2.7 \pm 1.9) \times 10^{-2}$	$(7.7 \pm 5.7) \times 10^{-3}$
$[\mu^+\pi^- \ \mu^+]$			
$B^{0} \to J/\psi \{ \to \mu^{+}\mu^{-} \} K^{*0} \{ \to K^{+}\pi^{-} \}$	$(6.0 \pm 0.2) \times 10^{-2}$	$(6.5 \pm 4.8) \times 10^{-4}$	$(5.1 \pm 3.2) \times 10^{-7}$
$B^0_s \to J/\psi \{ \to \mu^+ \mu^- \} \phi \{ \to K^+ K^- \}$	$(2.0 \pm 1.0) \times 10^{-2}$	$(5.5 \pm 1.0) \times 10^{-4}$	$(6.0 \pm 3.2) \times 10^{-5}$
All	$5.3 \times 10^{-4}$	$(1.2 \pm 0.2) \times 10^{-1}$	$(2.9 \pm 0.6) \times 10^{-2}$

Table 4.4: A summary of misidentified backgrounds for  $B^+ \to \pi^- \mu^+ \mu^+$ . The PID assignments required for each decay mode to be identified as  $B^+ \to \pi^- \mu^+ \mu^+$  are given in square brackets, with the order of the final state particles the same as in the original decay.

Firstly, the decay probability is assumed to be constant across phase-space. Secondly, the effect of the decay proceeding via a narrow  $n \to h^- \mu^+$  resonance, as would occur in the model described in Sect. 4.1, is assessed as a function of  $h^- \mu^+$  mass. For each value of  $h^- \mu^+$  mass, three distributions of  $M^2_{h^- \mu^+}$  are considered: a flat distribution, and linear gradients of +1 and -1. These represent the three possible extreme cases. The limit on the branching fraction is set using the  $M^2_{h^- \mu^+}$  distribution which gives the highest efficiency value, i.e the most conservative limit is taken.

The efficiencies are calculated from simulated  $B^+ \to J/\psi K^+$ ,  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$  events. Assuming a flat phase-space distribution for  $B^+ \to h^- \mu^+ \mu^+$ , the relative efficiency of  $B^+ \to K^- \mu^+ \mu^+$  ( $B^+ \to \pi^- \mu^+ \mu^+$ ) and  $B^+ \to J/\psi K^+$  was calculated to be  $(89.1 \pm 0.4 (\text{stat}) \pm 0.2 (\text{syst}))\%$  ( $(82.7 \pm 0.6 (\text{stat}) \pm 0.9 (\text{syst}))\%$ ). Assuming the decay is mediated exclusively by a narrow  $h^- \mu^+$  resonance, the relative efficiency of  $B^+ \to K^- \mu^+ \mu^+$  ( $B^+ \to \pi^- \mu^+ \mu^+$ ) and  $B^+ \to J/\psi K^+$  is show in Fig. 4.6. Taking Eqn. 4.2, together with the measured  $B^+ \to J/\psi K^+$  yield from Sect. 4.2.1, the  $B^+ \to J/\psi K^+$  branching fraction from Ref. [61] and the calculated efficiencies gives a single event sensitivity ( $\alpha$ ) of  $2.0 \times 10^{-8}$  for  $B^+ \to K^- \mu^+ \mu^+$ , and  $2.1 \times 10^{-8}$  for  $B^+ \to \pi^- \mu^+ \mu^+$ .

# 4.5 Limit setting

A limit on the branching fraction of the  $B^+ \to h^-\mu^+\mu^+$  decay is set by counting the number of observed events in the signal mass window (described in Sect. 4.2.1),  $N_{\text{obs.}}$ . This is compared to the expected number of background events,  $\mu_B$ , and a limit on the number of signal events in the sample,  $\mu_S$  determined. This limit on the number of signal events is then converted into a limit on the  $B^+ \to h^-\mu^+\mu^+$  branching fraction, using the single event sensitivity defined in Sect. 4.4. The number of events in the signal window,



Figure 4.5: The relative efficiency between  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to J/\psi K^+$  across phase space.



Figure 4.6: The relative efficiency between the signal and normalisation channels as a function of  $M_{h^-\mu^+}$ . The efficiencies at each value of  $M_{h^-\mu^+}$  are calculated for each of the three extreme phase space distributions detailed in the text.

N, follows a Poisson distribution with mean  $\mu = \mu_S + \mu_B$ , where  $\mu_S$  and  $\mu_B$  are the mean number of signal and background events respectively. The limit on S, at confidence level  $CL_{S+B}$  is given by the probability to observe  $N_{obs.}$  or fewer events in the signal mass window:

$$P_{S+B} = 1 - CL_{S+B} = P(N \le N_{obs.}|S+B).$$

The value of  $P(N \leq N_{obs.}|S + B)$  is calculated using toy datasets generated according to PDFs for the number of signal and background events. For each value of  $\mu_S$ ,  $P_{S+B}(N_{signal} \leq N_{obs.})$  is determined from the fraction of toy datasets containing  $N_{obs.}$  or fewer events in the signal mass window. The lowest value of  $\mu_S$  giving the desired  $CL_{S+B}$ is taken as the limit on the number of signal events. For each toy dataset,  $\mu_{bkg.}$  is drawn from a prior distribution encoding the uncertainty on the background expectation. The background is comprised of three components: combinatorial background, misidentified backgrounds without a  $\mu^+ \rightarrow h^+$  swap, and misidentified backgrounds with a  $\mu^+ \rightarrow h^+$
swap.

The uncertainty on the combinatorial background expectation has two sources: the number of events observed in the upper mass sideband, and the extrapolation of this yield into the signal window (as described in Sect. 4.2.2). The first source is treated as an auxiliary measurement following a Poisson distribution. For the extrapolation factor,  $\tau$ , a uniform prior in the range  $2 < \tau < 4.75$  is assumed.

The misidentified backgrounds with and without a  $\mu^+ \to h^+$  swap are treated separately due to the zero value for the  $\mu^+ \to h^+$  efficiencies. For the backgrounds without a  $\mu^+ \to h^+$  swap a Gaussian prior is taken for the background expectation. The dominant uncertainty on the backgrounds with a  $\mu^+ \to h^+$  swap is the  $\mu^+ \to h^+$  efficiency, which is treated as an auxiliary measurement following a Poisson distribution.

In order to include the systematic uncertainty on the single event sensitivity ( $\alpha$ , defined in Eqn. 4.2), for each toy dataset  $\mu_S$  is drawn from a Gaussian prior distribution, with a mean of unity, and a width corresponding to the uncertainty on the single event sensitivity.

# 4.6 Systematic uncertainties

#### 4.6.1 Background expectations

Two sources of uncertainty affect the expected yields for the misidentified backgrounds given in Sect. 4.3: the branching fractions of the decays in question, and the efficiency for these decays to pass the PID requirements. The uncertainties from the branching fractions are taken from Ref. [61]. To check the effect of the binning procedure used to calculate the efficiencies in Sect. 4.3, the efficiencies are recalculated in a finer binning scheme. This change in binning procedure produces a negligible change in the total PID efficiencies, and so no additional uncertainty is assigned.

#### 4.6.2 Single event sensitivity

#### Normalisation channel uncertainty

The  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$  branching fractions are measured relative to that of  $B^+ \to J/\psi K^+$ , and so the uncertainty on  $\mathcal{B}(B^+ \to J/\psi K^+)$  must be included in the determination of  $\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-)$ . The  $B^+ \to J/\psi K^+$  branching fraction is taken from Ref. [61], and has an uncertainty of 3.4%. The  $B^+ \to J/\psi K^+$  yield is determined from a fit to the  $M_{K^+\mu^+\mu^-}$  distribution, as described in Sect. 4.2.1. To assess the uncertainty due to the choice of PDFs, the fit is repeated with the Crystal Ball PDF used to model the  $B^+ \to J/\psi K^+$  signal changed to a Gaussian, and/or the polynomial used to model the combinatorial background changed to an exponential. The largest deviation between these four variations, 1.7%, is included as a systematic uncertainty.

#### **PID** performance

The statistical uncertainties on the efficiency of the PID requirements listed in Tab. 4.2 are included as a systematic uncertainty. The relative efficiency between  $B^+ \to h^- \mu^+ \mu^+$ and  $B^+ \to J/\psi K^+$  is repeatedly recalculated with the PID efficiencies varied according to their statistical uncertainties, in each of the kinematic bins described in Sect. 4.3. The width of the resulting distribution of relative efficiencies gives a systematic uncertainty of 0.1% for the  $B^+ \to K^- \mu^+ \mu^+$  final state, and an uncertainty of 0.8% for the  $B^+ \to \pi^- \mu^+ \mu^+$ final state. The uncertainty is larger for the  $B^+ \to \pi^- \mu^+ \mu^+$  final state due to the  $\pi^+ \to \pi^+$ efficiency not being shared with the normalisation channel, in contrast to the  $K^+ \to K^+$ efficiency in the case of  $B^+ \to K^- \mu^+ \mu^+$ .

#### **IP** resolution

The simulated samples generated for this analysis do not correctly reproduce the IP resolution in data. To assess the uncertainty arising from this mismodelling, the relative efficiencies between  $B^+ \rightarrow h^- \mu^+ \mu^+$  and  $B^+ \rightarrow J/\psi K^+$  are recalculated with the  $IP\chi^2$  requirement varied to 30 and 60 from the nominal 45. The largest variation, 0.2% is taken as a systematic uncertainty.

#### **Trigger efficiency**

The configuration of the trigger algorithms varied during the 2010 data taking period, whereas the simulated events are produced with the final, most restrictive, configuration. To assess the possible effects of this variation, the relative efficiencies are recalculated using an alternative, less restrictive trigger configuration. The resulting 0.1% variation is included as a systematic uncertainty.

#### Tracking efficiency

The tracking efficiency as a function of track momentum not being modelled correctly in the simulation is a potential source of uncertainty. The effects of such a mismodelling are estimated by weighting all tracks with a momentum below 5 GeV/c with a correction factor of 96%. This factor of 96% below 5 GeV/c is motivated by preliminary measurements of the relative tracking efficiency between data and simulated events. The relative efficiencies between  $B^+ \rightarrow h^- \mu^+ \mu^+$  and  $B^+ \rightarrow J/\psi K^+$  are recalculated with this reweighting, and the resulting 0.1% difference is included as a systematic uncertainty.

Source	$B^+ \rightarrow K^- \mu^+ \mu^+$	$B^+ \to \pi^- \mu^+ \mu^+$
$\mathcal{B}(B^+ \to J/\psi  K^+)$	3.4%	3.4%
$B^+ \rightarrow J/\psi K^+$ fit models	1.6%	1.6%
Simulation statistics	0.4%	0.6%
IP resolution	0.2%	0.2%
PID	0.1%	0.8%
Trigger efficiency	0.1%	0.1%
Tracking efficiency	0.1%	0.1%

Table 4.5: Sources of systematic error and their relative uncertainty on the single event sensitivity for  $B^+ \to K^- \mu^+ \mu^+$  and  $B^+ \to \pi^- \mu^+ \mu^+$ .



Figure 4.7: The invariant mass distribution of  $K^-\mu^+\mu^+$  (a) and  $\pi^-\mu^+\mu^+$  (b) events after the application of the selection criteria. No events in either decay mode survive the selection criteria



Figure 4.8: The 90% C.L limits set on the branching fractions of the  $B^+ \to K^- \mu^+ \mu^+$ (light-coloured line) and  $B^+ \to \pi^- \mu^+ \mu^+$  (dark-coloured line) decays, as a function of  $m_{h^-\mu^+}$ .

# 4.7 Results and conclusion

The  $M_{h^-\mu^+\mu^+}$  mass distribution for candidates passing the selection criteria are shown in Fig. 4.7. No  $B^+ \to K^-\mu^+\mu^+$  or  $B^+ \to \pi^-\mu^+\mu^+$  candidates are selected in the entire mass range. The limits on  $\mathcal{B}(B^+ \to K^-\mu^+\mu^+)$  and  $\mathcal{B}(B^+ \to \pi^-\mu^+\mu^+)$  are therefore both calculated for  $N_{\text{obs.}} = 0$ . This corresponds to branching fraction measurements of:

$$\mathcal{B}(B^+ \to K^- \mu^+ \mu^+) < 4.1 \times 10^{-8}$$
 at 90% C.L  
 $\mathcal{B}(B^+ \to \pi^- \mu^+ \mu^+) < 4.4 \times 10^{-8}$  at 90% C.L.

These improve the previous best limits by a factor 40 and 30, respectively [60]. The limits on  $\mathcal{B}(B^+ \to h^- \mu^+ \mu^+)$  if the decay proceeds entirely through a narrow  $h^+ \mu^-$  resonance are shown in Fig. 4.8 as a function of  $h^+ \mu^-$  mass. These limits remain at least five orders of magnitude above the theoretically favoured region of the  $\nu MSM$  model described in Sect. 4.1.

In the time after the publication of this result, the  $B^+ \to \pi^- \mu^+ \mu^+$  branching fraction limit has been lowered by an order of magnitude using the ~ 3 fb<sup>-1</sup> 2011 and 2012 LHCb dataset [63]. The LHCb upgrade may allow further improvements in this limit, but *B* decays will not access the theoretically favoured region of the  $\nu MSM$  in the foreseeable future. Recently, a dedicated experiment has been proposed to reach into the theoretically favoured mixing region, in the neutrino mass range accessible with *D* meson decays [64].

# Chapter 5

# First observation of $B^+ \rightarrow \pi^+ \mu^+ \mu^-$

## 5.1 Introduction

The ratio of CKM matrix elements  $|V_{td}|/|V_{ts}|$  has been measured in *B* mixing processes, where it is probed in box diagrams through the ratio of  $B^0$  and  $B_s^0$  mixing frequencies [65– 68]. The ratio of these matrix elements has also been measured using the ratio of branching fractions of  $b \rightarrow s\gamma$  and  $b \rightarrow d\gamma$  decays, where radiative penguin diagrams mediate the transition [69–71]. These measurements of  $|V_{td}|/|V_{ts}|$  are consistent, within the (dominant) ~10% uncertainty on the determination from radiative decays.

The decays  $b \to s\mu^+\mu^-$  and  $b \to d\mu^+\mu^-$  offer an alternative way of measuring  $|V_{\rm td}|/|V_{\rm ts}|$  which is sensitive to different classes of operators than the radiative decay modes [72]. These  $b \to (s, d)\mu^+\mu^-$  transitions are flavour-changing neutral current processes which, in the SM, are forbidden at tree level. In the SM, the branching fractions for  $b \to d\ell^+\ell^-$  transitions are suppressed relative to  $b \to s\ell^+\ell^-$  processes by the ratio  $|V_{\rm td}|^2/|V_{\rm ts}|^2$ . This suppression does not necessarily apply to models beyond the SM, and  $B^+ \to \pi^+\mu^+\mu^-$  decays may have a different sensitivity to the effect of new particles than  $B^+ \to K^+\mu^+\mu^-$  decays. In the SM, the ratio of branching fractions for these exclusive modes

$$R \equiv \mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-) / \mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)$$
(5.1)

is given by  $R = V^2 f^2$ , where  $V = |V_{td}|/|V_{ts}|$  and f is the ratio of the relevant form factors and Wilson coefficients, integrated over the relevant phase space. A difference between the measured value of R and  $V^2 f^2$  would indicate a deviation from the minimal flavour violation hypothesis [73, 74], and would rule out certain approximate flavour symmetry models [75].

No  $b \to d\ell^+\ell^-$  transitions have previously been measured, and the observation of the  $B^+ \to \pi^+\mu^+\mu^-$  decay would therefore be the first time such a process has been detected. The predicted SM branching fraction for  $B^+ \to \pi^+\mu^+\mu^-$  is  $(2.0 \pm 0.2) \times 10^{-8}$  [24], with the uncertainty mainly arising from the calculation of the  $B \to \pi$  form factors. The most stringent limit to date is  $\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-) < 6.9 \times 10^{-8}$  at 90% confidence level [76]. The analogous  $b \to s \ell^+ \ell^-$  decay,  $B^+ \to K^+ \mu^+ \mu^-$ , has been observed with a branching fraction measured relative to  $B^+ \to J/\psi K^+$  of  $(4.36 \pm 0.15 \pm 0.18) \times 10^{-7}$  [77], where the first error is statistical, the second systematic, and the third arises from the uncertainty on the  $B^+ \to J/\psi K^+$  branching fraction.

#### 5.1.1 Overview

This chapter describes the search for the  $B^+ \to \pi^+ \mu^+ \mu^-$  decay using proton-proton collision data, corresponding to an integrated luminosity of 1.0 fb<sup>-1</sup>, collected with the LHCb detector. The  $B^+ \to \pi^+ \mu^+ \mu^-$  branching fraction is measured with respect to that of  $B^+ \to J/\psi (\to \mu^+ \mu^-) K^+$  (hereafter denoted  $B^+ \to J/\psi K^+$ ), and the ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  branching fractions is also determined. One of the main backgrounds to  $B^+ \to \pi^+ \mu^+ \mu^-$  is  $B^+ \to K^+ \mu^+ \mu^-$ , with the kaon misidentified as a pion. The decay  $B^+ \to J/\psi K^+$  is used to constrain both the shape of the  $B^+ \to \pi^+ \mu^+ \mu^$ and  $B^+ \to K^+ \mu^+ \mu^-$  signal mass peaks and, in the  $B^+ \to \pi^+ \mu^+ \mu^-$  case, the invariant mass distribution of the misidentified  $B^+ \to K^+ \mu^+ \mu^-$  events.

Sources of background are described in Sect. 5.2, along with the candidate selection criteria designed to minimise these backgrounds. The modelling of the invariant mass lineshape of the misidentified  $B^+ \to K^+\mu^+\mu^-$  decay is described in Sect. 5.3. The fit used to extract the  $B^+ \to \pi^+\mu^+\mu^-$ ,  $B^+ \to K^+\mu^+\mu^-$  and  $B^+ \to J/\psi K^+$  yields is then described in Sect. 5.4. The computation of  $\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-)$  from the  $B^+ \to J/\psi K^+$  and  $B^+ \to \pi^+\mu^+\mu^-$  yields is described in Sect. 5.5. The systematic uncertainties affecting all of the measurements in this chapter are then described in Sect. 5.6, before these results are presented in Sect. 5.7. This measurement has been published in Ref. [2].

# 5.2 Event selection

The  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  candidates are selected by combining pairs of oppositely charged muons with a charged pion or kaon, with the individual particles and the resulting *B* candidate required to pass the set of preselection criteria that are given in Tab. 5.1. These preselection criteria serve to reduce the size of the dataset to a level the LHCb computing framework can accomodate.

The selection criteria are designed to reduce several backgrounds: resonant backgrounds, where the dimuon originates from a  $J/\psi$  decay; combinatorial backgrounds, where the particles selected do not originate from a single decay; peaking backgrounds, where a single decay is selected but with one or more particles misidentified; and partially reconstructed backgrounds, where one or more final-state particles from a *B* decay are not reconstructed. These backgrounds are each described below, along with the number

Selection	Criteria
$M_{h^+\mu^+\mu^-}$	$4900 < M_{h^+\mu^+\mu^-} < 7000 \mathrm{MeV}/c^2$
$M_{\mu^+\mu^-}$	$< 5050 \mathrm{MeV}/c^2$
$\mu^+ \chi^2_{IP}$	> 16
$h^+ \chi^2_{IP}$	> 9
$B^+ \chi^2_{IP}$	< 9
$\mu^+ \mu^-$ vertex $\chi^2/\text{ndof}$	< 9
$B^+$ vertex $\chi^2/ndof$	< 8
$B^+ \theta_P$	$< 8 \mathrm{mrad}$
$B^+$ flight distance $\chi^2$	> 100
$\mu^+ \text{ DLL}_{\mu\pi}$	> 0
$\mu^+$ is Muon	True

Table 5.1: The selection criteria used in the preselection.

of candidates expected to survive the selection criteria.

#### 5.2.1 Dimuon Mass Veto

In order to measure the non resonant processes  $B^+ \to h^+\mu^+\mu^-$ , it is necessary to veto the  $J/\psi$  and  $\psi(2S)$  resonances, which contribute both via genuine B decays  $(B^+ \to J/\psi h^+)$  and via the combination of a  $J/\psi$  with a hadron from another decay. Both of these backgrounds are clearly visible in a plot of dimuon mass against  $M_{K^+\mu^+\mu^-}$  mass, as shown in Fig. 5.1. The dimuon mass veto is illustrated in Fig. 5.2, and has a total width of 200 (150) MeV/ $c^2$  around the nominal  $J/\psi$  ( $\psi(2S)$ ) mass [8]. Candidates where the dimuon mass is shifted, due to mismeasurement or final state radiation, have a correlated shift in the  $h^+\mu^+\mu^-$  mass. The veto  $J/\psi$  and  $\psi(2S)$  veto therefore includes a component which shifts with  $M_{h^+\mu^+\mu^-}$  to exclude  $B^+ \to J/\psi h^+$  candidates with such a mismeasurement. The total width of the veto in dimuon mass remains constant with  $h^+\mu^+\mu^-$  mass. To first order, the amount of combinatorial background removed by the veto is constant across the  $M_{h^+\mu^+\mu^-}$  spectrum, and there is therefore no need to correct the combinatorial background level in a subsequent fit to the  $M_{h^+\mu^+\mu^-}$  spectrum. The distributions of dimuon mass veto, with the  $J/\psi$  and  $\psi(2S)$  resonances no longer giving a visible contribution.

#### 5.2.2 Combinatorial backgrounds

A BDT which employs the AdaBoost algorithm is used to separate signal candidates from the combinatorial background. The input variables to the BDT are listed in Tab. 5.2. In addition to the properties of the candidate B and daughters, the difference between the momenta of the two muons,  $\Delta P(\mu^+\mu^-)$ , and the primary vertex ndof are included. Two muons coming from different mother particles should have momenta less similar than those that come from the same mother particle. This is borne out by Fig. 5.4,



Figure 5.1: The dimuon and  $K^+\mu^+\mu^-$  mass for selected  $K^+\mu^+\mu^-$  events, before application of any mass vetoes.



Figure 5.2: The dimuon mass veto illustrated with randomly generated mass points.



Figure 5.3: The dimuon and  $K^+\mu^+\mu^-$  mass for selected  $K^+\mu^+\mu^-$  events, after application of the dimuon mass veto.



Figure 5.4: Normalised distributions of  $\Delta P(\mu^+\mu^-)$  for signal, shown in black, and sideband, shown in blue (medium grey) regions of  $B^+ \to J/\psi K^+$  data, and for simulated  $B^+ \to \pi^+ \mu^+ \mu^-$  events, shown in light grey).



Table 5.2: The variables used in the BDT.

which shows normalised plots of  $\Delta P(\mu^+\mu^-)$  for the sideband (5340 – 7000 MeV/ $c^2$ ) and signal (5240 – 5320 MeV/ $c^2$ ) region of  $B^+ \rightarrow J/\psi K^+$ , with the preselection requirements applied. This demonstrates that  $\Delta P(\mu^+\mu^-)$  has some separating power. The equivalent plot for the number of primary vertex degrees of freedom is shown in Fig. 5.5. The primary vertex ndof (equivalent to the number of tracks used to form the primary vertex) provides discrimination against non proton-proton interaction vertices (e.g other *B* decay vertices), identified as primary vertices.

The BDT is trained on a simulated  $B^+ \to \pi^+ \mu^+ \mu^-$  signal test sample, containing  $\sim 10^4$  reconstructed and selected events, and a background test sample taken from sidebands in the  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  invariant mass distributions.

The background test sample consists of 20% of the candidates with  $M_{\pi^+\mu^+\mu^-}$  or  $M_{K^+\mu^+\mu^-} > 5500 \text{ MeV}/c^2$ , and contains ~ 1000 reconstructed and selected events. This sample is not used for any of the subsequent analysis. A 50% portion of the background sample is used to train the BDT on, and a further 25% is used for optimising the BDT output cut chosen. The remaining 25% of the background sample is used neither to select



Figure 5.5: Normalised distributions of the number of primary vertex degrees of freedom for signal, shown in black and sideband, shown in light grey, regions of  $B^+ \to J/\psi K^+$  data.

the BDT output cut to apply, nor for BDT training, and therefore allowed an unbiased estimate of the performance of a chosen BDT output cut to be made.

Signal candidates are required to have a BDT output which exceeds a set value. The BDT output requirements are chosen by running toy experiments using the fit described in Sect. 5.4. Toy datasets are generated according to the expected yields for each source of candidates, the dataset fitted and the significance of the  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  signal determined. The significance is determined from the difference in the minimum log-likelihood between the signal-plus-background and background-only hypotheses. This procedure is repeated for different BDT requirements, as shown in Fig. 5.6. The median significance of these toy experiments is used as the figure of merit to determine the optimal BDT requirement.

For optimising the BDT requirement, the signal and background expectations are taken from the full data sample, using the cut-based selection which is detailed in Sect. 4.2. The signal expectation is calculated from the  $B^+ \to K^+ \mu^+ \mu^-$  yield observed in the data, scaled to the expected SM  $B^+ \to \pi^+ \mu^+ \mu^-$  branching fraction. Similarly, the misidentified  $B^+ \to K^+ \mu^+ \mu^-$  yield is taken from  $B^+ \to K^+ \mu^+ \mu^-$  in data, scaled by the efficiency to misidentify a kaon as a pion. The combinatorial background expectation is taken from the  $B^+ \to \pi^+ \mu^+ \mu^-$  upper mass sideband. These expectations are then scaled by the BDT efficiency relative to the cut-based selection, as measured on the background test samples described above.

The optimal requirement on the BDT output is determined to be > 0.325. This requirement reduces the expected combinatorial background from  $652 \pm 11$  to  $9 \pm 2$  candidates in a  $\pm 60 \text{ MeV}/c^2$  window around the nominal *B* mass, while retaining 68% of signal events. The uncertainties on these expectations are taken from an exponential fit to the upper mass sideband,  $M_{\pi^+\mu^+\mu^-}$  > 5500 MeV/ $c^2$ , extrapolated into the signal window. Assuming the SM branching fraction and the single event sensitivity (defined in Sect. 5.5) result-



Figure 5.6: The significance obtained from toy fits, based on different BDT and  $\text{DLL}_{\mu\pi}$  cuts.

ing from this requirement,  $21 \pm 3 B^+ \rightarrow \pi^+ \mu^+ \mu^-$  signal events are expected in the data sample. The BDT output distribution for simulated  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  events and for mass sideband candidates is shown in Fig. 5.7, with the BDT requirement chosen indicated.

#### 5.2.3 Peaking and partially reconstructed backgrounds

In addition to combinatorial backgrounds, other backgrounds consisting of a genuine, fully reconstructed  $B^+$  decay with one or more final state particles misidentified must be controlled. Such backgrounds have a measured mass which is shifted from the nominal B mass. These backgrounds are suppressed using PID requirements on the final state particles. The requirements are chosen using the same toy experiment procedure as is used to select the BDT requirement in Sect. 5.2.2.

The most significant such background is  $B^+ \to K^+ \mu^+ \mu^-$ , with the kaon misidentified as a pion. Due to the large  $B^+ \to K^+ \mu^+ \mu^-$  branching fraction relative to  $B^+ \to \pi^+ \mu^+ \mu^-$ , a  $B^+ \to K^+ \mu^+ \mu^-$  sample can be selected with negligible  $B^+ \to \pi^+ \mu^+ \mu^-$  contamination, using a strong kaon PID requirement. Scaling the  $B^+ \to K^+ \mu^+ \mu^-$  yield in this sample by the relative efficiency of the PID requirement used to isolate  $B^+ \to \pi^+ \mu^+ \mu^-$  and that used to isolate  $B^+ \to K^+ \mu^+ \mu^-$  gives an expectation for the residual  $B^+ \to K^+ \mu^+ \mu^$ yield of 6.2  $\pm$  0.3 candidates. The uncertainty on this expectation is derived from the statistical uncertainty on the  $B^+ \to K^+ \mu^+ \mu^-$  yield, and the uncertainty on the relative PID efficiency.

The only other decay found to give a significant peaking background in the search for  $B^+ \to \pi^+ \mu^+ \mu^-$  is  $B^+ \to \pi^+ \pi^+ \pi^-$ , where both a  $\pi^+$  and a  $\pi^-$  are misidentified as muons. For  $B^+ \to K^+ \mu^+ \mu^-$  decays, the only significant peaking background is  $B^+ \to K^+ \pi^+ \pi^-$ , which includes the contribution from  $B^+ \to D^0(\to K^+ \pi^-)\pi^+$ . The expected background levels from  $B^+ \to \pi^+ \pi^+ \pi^-$  ( $B^+ \to K^+ \pi^+ \pi^-$ ) decays are computed to



Figure 5.7: BDT output distribution for simulated  $B^+ \to \pi^+ \mu^+ \mu^-$  events (black solid line) and candidates taken from the mass sidebands in the data (red dotted line). Both distributions are normalised to unit area. The vertical line indicates cut value of 0.325, which maximises the expected  $B^+ \to \pi^+ \mu^+ \mu^-$  signal significance.

be  $0.39 \pm 0.04$  (1.56  $\pm 0.16$ ) residual background candidates, using simulated events. The dominant uncertainty on this expectation is the measured  $B^+ \rightarrow \pi^+ \pi^+ \pi^-$  branching fraction, which is taken from Ref. [8].

Backgrounds from B decays that have one or more final state particles which are not reconstructed have a mass below the nominal B mass, and do not extend into the signal window. However, in the  $B^+ \to \pi^+ \mu^+ \mu^-$  case, these backgrounds overlap with the misidentified  $B^+ \to K^+ \mu^+ \mu^-$  component described above, and must therefore be included in the fit. No expectations are calculated for these partially reconstructed backgrounds, and their yield is not constrained in the mass fit described in Sect. 5.4. In the  $B^+ \to K^+ \mu^+ \mu^-$  case, such partially reconstructed backgrounds are negligible in the relevant  $K^+ \mu^+ \mu^-$  mass range.

#### 5.2.4 Control channels

The  $B^+ \to J/\psi K^+$  and  $B^+ \to K^+ \mu^+ \mu^-$  decays are used to normalise the number of  $B^+ \to \pi^+ \mu^+ \mu^-$  candidates observed, and to constrain several invariant mass distributions. Candidates are isolated by replacing the pion PID criteria with a requirement to select kaons. In addition, instead of the dimuon mass vetoes described above, the  $B^+ \to J/\psi K^+$  candidates are required to have dimuon mass within  $\pm 50 \text{ MeV}/c^2$  of the nominal  $J/\psi$  mass (the  $J/\psi$  mass resolution is 14.5 MeV/ $c^2$ ). The remainder of the selection is the same as that used to isolate  $B^+ \to \pi^+ \mu^+ \mu^-$  candidates. This minimises the systematic uncertainty on the ratio of branching fractions, although the resulting selection is considerably tighter than that which would give the lowest statistical uncertainty on the  $B^+ \to K^+ \mu^+ \mu^-$  event yield. The  $B^+ \to J/\psi \pi^+$  candidates, which are discussed below, are selected using the same

BDT, the pion PID criteria, and the above window on the dimuon invariant mass. There is no significant peaking background for  $B^+ \to J/\psi K^+$  decays. For  $B^+ \to J/\psi \pi^+$  decays the only significant peaking background is misidentified  $B^+ \to J/\psi K^+$  events. After all selection critera are imposed, the rate of events containing more than one reconstructed candidate is 1 in ~20,000 for  $B^+ \to J/\psi K^+$ . No restriction is therefore placed on the number of candidates per event.

# 5.3 Misidentified $B^+ \rightarrow K^+ \mu^+ \mu^-$ lineshape

The shift in invariant mass when reconstructing the  $M_{\pi^+\mu^+\mu^-}$  distribution of  $B^+ \to K^+\mu^+\mu^-$  is given by:

$$\Delta(M^2) = \left[ \left( E_D + \sqrt{P_h^2 + m_K^2} \right)^2 - P_B^2 \right] - \left[ \left( E_D + \sqrt{P_h^2 + m_\pi^2} \right)^2 - P_B^2 \right]$$
(5.2)

$$= m_K^2 - m_\pi^2 + 2E_D \left(\sqrt{P_h^2 + m_K^2} - \sqrt{P_h^2 + m_\pi^2}\right), \qquad (5.3)$$

where  $E_D$  is the energy of the dimuon system,  $P_B$  is the momentum of the *B* candidate, and  $P_h$  is the momentum of the hadron. The  $M_{\pi^+\mu^+\mu^-}$  distributions of  $B^+ \to K^+\mu^+\mu^-$  and  $B^+ \to J/\psi K^+$  will therefore differ due to differing dimuon energy and hadron momentum spectra. This is shown in Fig. 5.8(a), for simulated  $B^+ \to K^+\mu^+\mu^-$  and  $B^+ \to J/\psi K^+$ samples. The difference in  $M_{\pi^+\mu^+\mu^-}$  lineshape is corrected for by reweighting  $B^+ \to J/\psi K^+$ candidates to match  $B^+ \to K^+\mu^+\mu^-$  in the two dimensional dimuon energy and hadron momentum distribution. The samples are shown again in Fig. 5.8(b) after this reweighting, and are then in good agreement.

The PID requirements used in the selection have a momentum dependent efficiency, with a different dependence between the requirement used to isolate  $B^+ \to \pi^+ \mu^+ \mu^-$ , and that used to isolate  $B^+ \to J/\psi K^+$ . This will introduce a difference in the hadron momentum spectra between the two samples, and so a difference in the  $M_{\pi^+\mu^+\mu^-}$  distributions. In order to correct for this effect, the  $B^+ \to J/\psi K^+$  candidates are reweighted according to the PID efficiencies derived from data, as described in Sect. 2.2. This adjusts the  $B^+ \to J/\psi K^+$  invariant mass distribution to remove the effect of the kaon PID requirement that is used to isolate  $B^+ \to J/\psi K^+$  candidates, and to reproduce the effect of the pion PID requirement that is used to isolate  $B^+ \to \pi^+ \mu^+ \mu^-$  candidates.

## 5.4 Signal yield determination

The  $B^+ \to \pi^+ \mu^+ \mu^-$ ,  $B^+ \to K^+ \mu^+ \mu^-$  and  $B^+ \to J/\psi K^+$  yields are determined from a simultaneous unbinned maximum likelihood fit to four invariant mass distributions taken



Figure 5.8: The mass distribution of selected simulated  $B^+ \to J/\psi K^+$  events, shown in grey, and  $B^+ \to K^+ \mu^+ \mu^-$  events, shown in black. In (b),  $B^+ \to J/\psi K^+$  has been reweighted to match the  $B^+ \to K^+ \mu^+ \mu^-$  distributions in hadron momentum and dimuon mass.

from the data, which contain:

- 1. Reconstructed  $B^+ \rightarrow J/\psi K^+$  candidates;
- 2. Reconstructed  $B^+ \rightarrow J/\psi K^+$  candidates, with the kaon attributed to have the pion mass;
- 3. Reconstructed  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  candidates; and
- 4. Reconstructed  $B^+ \rightarrow K^+ \mu^+ \mu^-$  candidates.

All fit parameters are freely floating unless otherwise stated. The signal PDFs for the  $B^+ \to \pi^+ \mu^+ \mu^-$ ,  $B^+ \to K^+ \mu^+ \mu^-$ , and  $B^+ \to J/\psi K^+$  decay modes are modelled with the sum of two Gaussian functions. The PDFs for all of these decay modes share the same mean, widths and fraction of the total PDF between the two Gaussians. The  $B^+ \to \pi^+ \mu^+ \mu^-$  PDF is adjusted for the difference between the widths of the  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to J/\psi K^+$  distributions, which is observed to be at the percent level in simulation. This is implemented by multiplying the shared widths by a scale factor in the  $B^+ \to \pi^+ \mu^+ \mu^-$  case, which is constrained in the fit. The peaking backgrounds described in Sect. 5.2.3 are taken into account in the fit by including PDFs with shapes determined from simulation, and with their yields constrained to their expectations. The combinatorial backgrounds are modelled with a single exponential PDF, with the exponent allowed to vary independently for each distribution. The partially reconstructed candidates are modelled using a PDF consisting of an exponential distribution cut-off at a threshold mass, with the transition smeared by the experimental resolution. The shape parameters are again allowed to vary independently for each distribution. The misidentified  $B^+ \to J/\psi K^+$  candidates are modelled with a Crystal Ball



Figure 5.9: Invariant mass distribution for  $B^+ \to J/\psi K^+$  candidates under the (a)  $K^+\mu^+\mu^-$  and (b)  $\pi^+\mu^+\mu^-$  mass hypotheses with the fit projections overlaid. In the legend, "part. reco" refers to partially reconstructed background. The fit models are described in the text.

function [78], as it describes the shape well. In all cases the PDFs are chosen empirically, without reference to a formal measure of goodness of fit. In order to describe the relevant background components for  $B^+ \to \pi^+ \mu^+ \mu^-$ , the fit is performed in the mass range  $4900 < M_{\pi^+\mu^+\mu^-} < 7000 \text{ MeV}/c^2$ . To avoid fitting the partially reconstructed background for  $B^+ \to K^+ \mu^+ \mu^-$ , which is irrelevant for the analysis, the fit is performed in the mass range  $5170 < M_{K^+\mu^+\mu^-} < 7000 \text{ MeV}/c^2$ .

# 5.4.1 Reconstructed $B^+ \rightarrow J/\psi K^+$ candidates

The reconstructed  $B^+ \to J/\psi K^+$  candidates are shown in the  $M_{K^+\mu^+\mu^-}$  distribution in Fig. 5.9(a). The fitted  $B^+ \to J/\psi K^+$  yield is 106,230 ± 330. This large event yield determines the lineshape for the  $B^+ \to \pi^+\mu^+\mu^-$  and  $B^+ \to K^+\mu^+\mu^-$  signal distributions, and provides the normalisation for the  $B^+ \to \pi^+\mu^+\mu^-$  branching fraction.

# 5.4.2 Reconstructed $B^+ \rightarrow J/\psi K^+$ candidates with the pion mass hypothesis

The  $B^+ \to J/\psi K^+$  candidates reconstructed under the pion mass hypothesis provide the lineshape for the misidentified  $B^+ \to K^+\mu^+\mu^-$  candidates that are a background to the  $B^+ \to \pi^+\mu^+\mu^-$  signal. The equivalent background from  $B^+ \to \pi^+\mu^+\mu^-$  in the  $B^+ \to K^+\mu^+\mu^-$  sample is negligible.

The  $M_{\pi^+\mu^+\mu^-}$  distribution after the weighting procedure described in Sect. 5.3 has been applied is shown in Fig. 5.9(b).



Figure 5.10: Invariant mass distribution of  $B^+ \to \pi^+ \mu^+ \mu^-$  candidates with the fit projection overlaid (a) in the full mass range and (b) in the region around the *B* mass. In the legend, "part. reco." and "combinatorial" refer to partially reconstructed and combinatorial backgrounds respectively.

# 5.4.3 Reconstructed $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates

The yield of misidentified  $B^+ \to K^+ \mu^+ \mu^-$  candidates in the  $B^+ \to \pi^+ \mu^+ \mu^-$  invariant mass distribution is constrained to the expectation of 6.2 ± 0.3 candidates computed in Sect. 5.2.3. Performing the fit without this constraint gives a yield of 5.6 ± 6.4 misidentified  $B^+ \to K^+ \mu^+ \mu^-$  candidates. The yields for the peaking background components are constrained to the expectations given in Sect. 5.2.3. For the  $B^+ \to K^+ K^- K^+$  and  $B^+ \to \pi^+ \pi^- \pi^+$  backgrounds, the signal PDF shape parameters are fixed. For both the  $M_{\pi^+\mu^+\mu^-}$  and  $M_{K^+\mu^+\mu^-}$  distributions, the exponential PDF used to model the combinatorial background has a step in the normalisation at 5500 MeV/ $c^2$  to account for the data used for training the BDT.

The  $M_{\pi^+\mu^+\mu^-}$  and  $M_{K^+\mu^+\mu^-}$  distributions are shown in Figs 5.10 and 5.11, respectively. The fit gives a  $B^+ \to \pi^+\mu^+\mu^-$  signal yield of 25.3  $^{+6.7}_{-6.4}$  candidates, and a  $B^+ \to K^+\mu^+\mu^-$  signal yield of 553  $^{+24}_{-25}$  candidates.

#### 5.4.4 Cross check of the fit procedure

The fit procedure was cross-checked on  $B^+ \to J/\psi \pi^+$  decays, accounting for the background from  $B^+ \to J/\psi K^+$  decays. The resulting fit is shown in Fig. 5.12. The shape of the combined  $B^+ \to J/\psi \pi^+$  and  $B^+ \to J/\psi K^+$  mass distribution is well reproduced. The  $B^+ \to J/\psi K^+$  yield is not constrained in this fit. The fitted yield of 1024 ± 61  $B^+ \to J/\psi K^+$  candidates is consistent with the expectation of 958 ± 31 candidates, where the errors on both yields are statistical only. This expectation is again computed by weighting the  $B^+ \to J/\psi K^+$  candidates, which are isolated using a kaon PID requirement, according to the PID efficiency derived from  $D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$  events.



Figure 5.11: Invariant mass distribution of  $B^+ \to K^+ \mu^+ \mu^-$  candidates with the fit projection overlaid (a) in the full mass range and (b) in the region around the nominal B mass. In the legend, "combinatorial" refers to the combinatorial background.



Figure 5.12: Invariant mass distribution of  $B^+ \rightarrow J/\psi \pi^+$  candidates with the fit projection overlaid. In the legend, "part. reco." and "combinatorial" refer to partially reconstructed and combinatorial backgrounds respectively. The fit model is described in the text.

### 5.5 Normalisation

The  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  branching fraction is given by

$$\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-) = \frac{\mathcal{B}(B^+ \to J/\psi K^+)}{N_{B^+ \to J/\psi K^+}} \cdot \frac{\epsilon_{B^+ \to J/\psi K^+}}{\epsilon_{B^+ \to \pi^+ \mu^+ \mu^-}} \cdot N_{B^+ \to \pi^+ \mu^+ \mu^-}, \qquad (5.4)$$

$$= \alpha \cdot N_{B^+ \to \pi^+ \mu^+ \mu^-}, \qquad (5.5)$$

where  $\mathcal{B}(X)$ ,  $N_X$  and  $\epsilon_X$  are the branching fraction, the number of events and the total efficiency, respectively, for decay mode X, and  $\alpha$  is the single event sensitivity.

#### 5.5.1 Reconstruction and selection efficiencies

The total efficiency can be divided into the following terms:

- $\epsilon_{reco}$  is the efficiency to reconstruct the tracks and decay vertex, including the geometric acceptance of the LHCb detector. It is calculated as the ratio between the number of decays generated and the number of decays reconstructed;
- $\epsilon_{presel|reco}$  is the proportion of reconstructed events passing the preselection criteria detailed in Sect. 5.2;
- $\epsilon_{sel|presel}$  is the proportion of preselected events passing the selection criteria detailed in Sect. 5.2;
- $\epsilon_{trig|sel}$  is the efficiency of the trigger requirements on events passing the selection criteria.

The ratio of efficiencies at each level between  $B^+ \to J/\psi K^+$  and  $B^+ \to \pi^+ \mu^+ \mu^-$  is given in Tab. 5.3. The difference in selection efficiencies between  $B^+ \to J/\psi K^+$  and  $B^+ \to \pi^+ \mu^+ \mu^-$  comes largely from the dimuon mass veto applied to  $B^+ \to \pi^+ \mu^+ \mu^-$ , and from the different hadron PID cuts applied. Removing the veto and PID cuts, the ratio of selection efficiencies ( $\epsilon_{sel|presel}$ ) becomes 1.027  $\pm$  0.006. The residual difference arises due to the difference in kinematics between  $B^+ \to J/\psi K^+$  and  $B^+ \to \pi^+ \mu^+ \mu^-$ .

Taking Eqn. 5.5 together with the measured  $B^+ \rightarrow J/\psi K^+$  yield from Sect. 5.4.1, the  $B^+ \rightarrow J/\psi K^+$  branching fraction from Ref. [8] and  $\epsilon_{B^+ \rightarrow J/\psi K^+}/\epsilon_{B^+ \rightarrow \pi^+ \mu^+ \mu^-} = 1.60 \pm 0.01$  gives a single event sensitivity of

$$\alpha = (9.34 \pm 0.03 \ (stat.) \pm 0.66 \ (syst.)) \times 10^{-10}.$$
(5.6)

The ratio of efficiencies at each level between  $B^+ \to K^+ \mu^+ \mu^-$  and  $B^+ \to \pi^+ \mu^+ \mu^$ is given in Tab. 5.4. The only large difference in efficiencies arises from the difference between kaon and pion PID, which appears in  $\epsilon_{sel|presel}$ . Removing the PID cuts, the ratio of selection efficiencies ( $\epsilon_{sel|presel}$ ) becomes 1.027 ± 0.005.

	$\epsilon_{reco}$	$\epsilon_{presel reco}$	$\epsilon_{sel presel}$	$\epsilon_{trig sel}$	$\epsilon_{tot}$
$\frac{B^+ \to J/\psi K^+}{B^+ \to \pi^+ \mu^+ \mu^-}$	$1.008 \pm 0.002$	$1.018 \pm 0.004$	$1.579 \pm 0.013$	$1.020 \pm 0.011$	$1.653 \pm 0.014$

Table 5.3: Values for the efficiencies shown in Eqn. (5.5), calculated using simulated events.

	$\epsilon_{reco}$	$\epsilon_{presel reco}$	$\epsilon_{sel presel}$	$\epsilon_{trig sel}$	$\epsilon_{tot}$
$\frac{B^+ \to K^+ \mu^+ \mu^-}{B^+ \to \pi^+ \mu^+ \mu^-}$	$1.001 \pm 0.001$	$1.001 \pm 0.003$	$1.169 \pm 0.008$	$0.988 \pm 0.009$	$1.167\pm0.008$

Table 5.4: Values for the efficiencies shown in Eqn. (5.5) for  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$ , calculated using simulated events.

#### 5.5.2 Change of form factors

The  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  simulation samples are generated using the form factors from Ref. [23], which have been superseded by improved calculations. The simulation is therefore corrected to match the dimuon mass spectrum produced by the form factors in Ref. [79]. This is implemented by reweighting the  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  samples according to the difference in generator-level dimuon mass spectra between the two sets of form factors. There is no other significant change in kinematics. After this reweighting, the ratio of  $B^+ \to J/\psi K^+$  and  $B^+ \to \pi^+ \mu^+ \mu^-$  efficiencies is 1.60  $\pm$  0.01, and the ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  to  $B^+ \to K^+ \mu^+ \mu^-$  efficiencies is 1.14  $\pm$  0.01. The  $B^+ \to \pi^+ \mu^+ \mu^-$  total efficiency is changed by 3%, while the ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  to  $B^+ \to K^+ \mu^+ \mu^-$  branching fraction and ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  branching fractions are both measured using the efficiencies calculated from these reweighted samples.

# 5.6 Systematics

There are two kinds of systematics which are considered: those affecting the determination of the  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  signal yields; and those affecting only the normalisation. Systematics which affect the yield extraction are included as constraints in the fit, and their effect is therefore included in the likelihood used to determine the  $B^+ \to \pi^+ \mu^+ \mu^-$  significance. Systematics which affect only the normalisation are summed in quadrature, and included as an uncertainty on the measured branching fraction.

#### 5.6.1 Fit uncertainties

In the simultaneous fit, the mass shapes are constrained by control modes, and the fit uncertainty on these shapes is propagated directly to the uncertainty on the signal yield. Systematics uncertainties on the mass shape differences between  $B^+ \to J/\psi K^+$  and  $B^+ \to K^+ \mu^+ \mu^-$  in the  $M_{\pi^+\mu^+\mu^-}$  fit are included as constraints on scale factors multiplying the misidentified  $B^+ \to K^+ \mu^+ \mu^-$  mass shape parameters. In addition, the yields of the misidentified  $B^+ \to K^+ \mu^+ \mu^-$  component, and those of the peaking background components, are constrained to their expectations, with the relevant uncertainties implemented as a Gaussian constraint. Uncertainties arising from the peaking background mass shapes are negligible.

#### $B^+ \rightarrow K^+ \mu^+ \mu^-$ shape uncertainties

Uncertainties in the shape parameters for the misidentified  $B^+ \to K^+ \mu^+ \mu^-$  PDF in the fit arising from the correction procedure described in Sect. 5.3 are taken into account by including Gaussian constraints on the values of the shape parameters. Three sources of uncertainty are considered: the statistical uncertainty on the  $B^+ \to K^+ \mu^+ \mu^-$  dimuon mass spectrum in data, the statistical uncertainty in the simulated  $B^+ \to K^+ \mu^+ \mu^-$  and  $B^+ \to J/\psi K^+$  samples, and the statistical uncertainty in the PID efficiencies. The fit to the  $B^+ \to K^+ \mu^+ \mu^-$  sample is repeated with the  $B^+ \to J/\psi K^+$  candidate weights described in Sect. 5.3 varied according to these three uncertainties, and the resulting distributions of mass shape parameters are fitted with Gaussian PDFs, which are then used as constraints in the fit.

#### Signal shape uncertainties

There is a slight width difference between  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$ . Using simulated  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  events, this difference is measured to be  $1.05 \pm 0.01$ , where the uncertainty is due to limited simulation statistics only. The width of the  $B^+ \to \pi^+ \mu^+ \mu^-$  signal PDF in the fit is scaled upwards by this factor, with the uncertainty on this factor included as a constraint.

#### Fit uncertainties summary

The uncertainty on the  $B^+ \to \pi^+ \mu^+ \mu^-$  yield determined with the fit takes all of the uncertainties described above into account, and they are therefore included in the statistical rather than the systematic uncertainty. These uncertainties affect the  $B^+ \to \pi^+ \mu^+ \mu^-$  yield at a level below one percent. None of these effects give rise to any significant uncertainty for the  $B^+ \to K^+ \mu^+ \mu^-$  decay.

#### 5.6.2 Normalisation uncertainties

Uncertainties on the two efficiency ratios  $\epsilon_{B^+\to J/\psi K^+}/\epsilon_{B^+\to\pi^+\mu^+\mu^-}$  and  $\epsilon_{B^+\to K^+\mu^+\mu^-}/\epsilon_{B^+\to\pi^+\mu^+\mu^-}$  affect the conversion of the  $B^+\to\pi^+\mu^+\mu^-$  yield into a branching fraction, and the measurement of the ratio of branching fractions, R, both described in Sect. 5.5. The size of the systematic uncertainties described below are listed in Tab. 5.5.

#### Normalisation channel uncertainty

The  $B^+ \to \pi^+ \mu^+ \mu^-$  branching fraction is measured relative to that of  $B^+ \to J/\psi K^+$ , and so the uncertainty on  $\mathcal{B}(B^+ \to J/\psi K^+)$  must be included in  $\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-)$ . The  $B^+ \to J/\psi K^+$  branching fraction is taken from Ref. [8], along with an uncertainty of 3.5%.

#### **Trigger efficiency**

In order to estimate the systematic uncertainty on the relative trigger efficiency between signal and normalisation channels, the trigger efficiency for  $B^+ \rightarrow J/\psi K^+$  is measured in both data and in simulated events using the method in Ref. [48]. The measured values of the trigger efficiency are 87.0 ± 2.6 % in data, and 85.8 ± 1.8 % in simulated events . The difference is taken as a systematic, giving a relative uncertainty of 1.4 %.

#### **IP** resolution

The simulated samples generated for this analysis do not correctly reproduce the IP resolution in data. This effect is corrected by a tool which smears the position of the detector hits in each track in a manner tuned to reproduce the IP resolution seen in data. The nominal efficiencies are calculated with this smearing applied. To estimate the uncertainty arising from this smearing, the efficiencies for  $B^+ \rightarrow J/\psi K^+$  and  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  are both recalculated without the smearing applied. The relative efficiencies change by 0.006%, and this difference is included as a systematic uncertainty.



Figure 5.13: Comparisons between data and simulated  $B^+ \rightarrow J/\psi K^+$  events for the BDT output. The smeared, shown in blue (medium grey) and unsmeared, shown in light grey simulated events are compared to data, shown in black.

#### Tracking efficiency

The efficiency for a track being reconstructed varies with the kinematics of the track, and is not perfectly reproduced by the simulation. The tracking efficiency has previously been measured, and the ratio between data and simulated efficiencies are known as a function of momentum and pseudorapidity. Reweighting both the simulated  $B^+ \rightarrow J/\psi K^+$  and  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  samples by this ratio, the relative efficiency changes by 0.28%, which is included as a systematic uncertainty.

#### **Data-simulation differences**

The selection efficiencies are calculated from a simulated sample, and so rely on the accuracy of the simulation. To check the quality of this simulation, all reconstructed quantities relevant to the selection are compared between simulation and data in the control channel  $B^+ \rightarrow J/\psi K^+$ . One dimensional comparisons for each variable are shown in Figs. A.1-A.6 in Appendix A, with the combinatorial background distribution estimated from the mass sidebands and subtracted. Similarly, the BDT output distribution is compared in Fig. 5.13.

Disagreements in the BDT input variables are corrected by sequential one-dimensional reweighting, the results of which are shown in Fig. A.7-A.12 for the BDT input variables and Fig. 5.14 for the BDT output. After the reweighting procedure, the BDT output distributions in data and simulated  $B^+ \rightarrow J/\psi K^+$  events is in good agreement, indicating that the correlations between variables are also sufficiently well reproduced. This same reweighting is also applied to the simulated  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  sample, and the relative selection efficiency recalculated. The ratio of selection efficiencies shifts by 0.27% with this reweighting applied to  $B^+ \rightarrow J/\psi K^+$  and  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ , and this difference is included as a systematic uncertainty.



Figure 5.14: Comparisons between the BDT output distributions for  $B^+ \to J/\psi K^+$  candidates selected from the data, shown in black, and from simulated events after reweighting, shown in grey.

#### **PID** performance

To check the efficiency of the  $\text{DLL}_K$  PID requirements, the ratio of  $B^+ \to J/\psi K^+$  to  $B^+ \to J/\psi \pi^+$  is measured, using the fit described in section 5.4.4, for different  $\text{DLL}_K$  requirements. The fitted ratio is corrected for the efficiency of the PID cut applied to the hadron, and the corrected yield compared to the nominal value. This procedure is performed for three sets of PID requirements used to isolate  $B^+ \to J/\psi K^+$  and  $B^+ \to J/\psi \pi^+$ : the nominal requirements, tighter requirements, and looser requirements.

The largest difference between any of these three measurements of the ratio of  $B^+ \to J/\psi K^+$  to  $B^+ \to J/\psi \pi^+$ , 1.1%, is taken as a systematic uncertainty.

#### Model dependence

The  $B^+ \to \pi^+ \mu^+ \mu^-$  efficiency varies with dimuon mass. If the true  $B^+ \to \pi^+ \mu^+ \mu^-$  dimuon mass distribution varies from that generated in the simulation, the  $B^+ \to \pi^+ \mu^+ \mu^-$  efficiency would therefore not be correct. To check this, the  $B^+ \to \pi^+ \mu^+ \mu^-$  efficiency is recalculated averaged over bins of dimuon mass. The  $B^+ \to \pi^+ \mu^+ \mu^-$  dimuon mass spectrum is extracted from data, and the average efficiency calculated. The background dimuon mass distribution is taken from the region 5500 – 6000 MeV/ $c^2$ , scaled to match the background expectation beneath the  $B^+ \to \pi^+ \mu^+ \mu^-$  peak, and subtracted from the  $B^+ \to \pi^+ \mu^+ \mu^-$  dimuon mass spectrum in data, the  $B^+ \to \pi^+ \mu^+ \mu^-$  efficiency relative to  $B^+ \to J/\psi K^+$  is  $\frac{\epsilon_{B^+ \to J/\psi K^+}}{\epsilon_{B^+ \to \pi^+ \mu^+ \mu^-}} = 1.345 \pm 0.420$ , consistent with the simulated value of  $1.224 \pm 0.005$ . The same check is made for  $B^+ \to K^+ \mu^+ \mu^-$ , with the average efficiency in data relative to  $B^+ \to \pi^+ \mu^+ \mu^-$ ,  $\frac{\epsilon_{B^+ \to \pi^+ \mu^+ \mu^-}}{\epsilon_{B^+ \to \pi^+ \mu^+ \mu^-}} = 1.105 \pm 0.049$ , consistent with the average of 1.102  $\pm 0.003$  from the simulation.

The change in form factor models described in Sect. 5.5.2 results in a 3.0% difference in the  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  branching fraction, and a 1.7% difference in the ratio of  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ 

Systematic	Uncertainty (%)
Branching fraction uncertainty	3.5
Trigger efficiency	1.4
IP smearing	0.006
Tracking efficiency	0.28
Data-simulation differences	0.27
PID performance	1.1
Simulation statistics	0.7
Form factors	3.0
Total	5.0

Table 5.5: Summary of normalisation systematics.

Systematic	Uncertainty $(\%)$
Trigger efficiency	1.4
IP smearing	0.006
Tracking efficiency	0.28
Data-simulation differences	0.27
PID performance	1.1
Simulation statistics	0.7
Form factors	1.7
Total	2.6

Table 5.6: Summary of systematics in normalising to  $B^+ \to K^+ \mu^+ \mu^-$ .

to  $B^+ \to K^+ \mu^+ \mu^-$ . These differences are taken as systematic uncertainties.

# 5.6.3 Normalisation to $B^+ \rightarrow K^+ \mu^+ \mu^-$

The same uncertainties which apply to normalising  $B^+ \to \pi^+ \mu^+ \mu^-$  to  $B^+ \to J/\psi K^+$ also apply to normalising  $B^+ \to \pi^+ \mu^+ \mu^-$  to  $B^+ \to K^+ \mu^+ \mu^-$ , with the exception of the uncertainty from the measured  $B^+ \to J/\psi K^+$  branching fraction. These uncertainties are summarised in Tab. 5.6.

# 5.7 Results and conclusion

The statistical significance of the  $B^+ \to \pi^+ \mu^+ \mu^-$  signal observed in Fig. 5.10 is computed from the difference in the minimum log-likelihood between the signal-plus-background and background-only hypotheses. Both the statistical and systematic uncertainties on the shape parameters (which affect the significance) are taken into account. The fitted yield corresponds to an observation of the  $B^+ \to \pi^+ \mu^+ \mu^-$  decay with 5.2  $\sigma$  significance. This is the first observation of a  $b \to d\ell^+\ell^-$  transition. Normalising the observed signal to the  $B^+ \to J/\psi K^+$  decay, using the single event sensitivity given in Sect. 5.5, the branching fraction of the  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  decay is measured to be

$$\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-) = (2.3 \pm 0.6 \text{ (stat)} \pm 0.1 \text{ (syst)}) \times 10^{-8},$$

where the first uncertainty is statistical and the second systematic. This is compatible with the SM expectation of  $(2.0 \pm 0.2) \times 10^{-8}$  [24]. Given the agreement between the present measurement and the SM prediction, contributions from physics beyond the SM can only modify the  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  branching fraction at the 10% level. A significant improvement in the precision of both the experimental measurements and the theoretical prediction will therefore be required to resolve any new physics contributions.

Taking the measured  $B^+ \to K^+ \mu^+ \mu^-$  yield and  $\epsilon_{B^+ \to K^+ \mu^+ \mu^-}/\epsilon_{B^+ \to \pi^+ \mu^+ \mu^-}$ , the ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  branching fractions is measured to be

$$\frac{\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-)}{\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)} = 0.053 \pm 0.014 \text{ (stat)} \pm 0.001 \text{ (syst)}.$$

In order to extract  $|V_{td}|/|V_{ts}|$  from this ratio of branching fractions, the SM expectation for the ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  branching fractions is calculated using the EVTGEN package [80], which implements the calculation in Ref. [23]. This calculation has been updated with the expressions for Wilson coefficients and power corrections from Ref. [81], and formulae for the  $q^2$  dependence of these coefficients from Refs. [82, 83]. Using this calculation, and form factors taken from Ref. [79] ("set II"), the integrated ratio of form factors and Wilson coefficients is determined to be f = 0.87. Neglecting theoretical uncertainties, the measured ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^$ branching fractions then gives

$$|V_{\rm td}|/|V_{\rm ts}| = \frac{1}{f} \sqrt{\frac{\mathcal{B}(B^+ \to \pi^+ \mu^+ \mu^-)}{\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)}} = 0.266 \pm 0.035 \text{ (stat)} \pm 0.003 \text{ (syst)},$$

which is compatible with previous determinations [68–71]. An additional uncertainty will arise from the knowledge of the form factors. As an estimate of the scale of this uncertainty, the "set IV" parameters available in Ref. [79] change the value of  $|V_{\rm td}|/|V_{\rm ts}|$  by 5.1%. This estimate is unlikely to cover a one sigma range on the form factor uncertainty, and does not take into account additional sources of uncertainty beyond the form factors (e.g potential contributions via annihilation diagrams). A full theoretical calculation taking into account such additional uncertainties, which also accurately determines the uncertainty on the ratio of form factors, would allow a determination of  $|V_{\rm td}|/|V_{\rm ts}|$  with comparable precision to that from radiative penguin decays.

# Chapter 6

# Sensitivity study for $B^0 \rightarrow D^{*-} \tau^+ \nu$

# 6.1 Introduction

In the SM, the ratio  $R_{D^{(*)}} = \mathcal{B}(B \to D^{(*)}\tau^+\nu) / \mathcal{B}(B \to D^{(*)}\mu^+\nu)$  differs from unity only due to the difference between the muon and tau invariant masses. The larger tau mass results in a reduced phase-space available to the  $B \to D^{(*)}\tau^+\nu$  decay, and this decay having a greater sensitivity to a helicity-suppressed form factor. The difference in phase-space and the effect of the form factors are theoretically well predicted, as described in Sect. 2.4.4, resulting in precise SM expectations of  $R_D = 0.296 \pm 0.016$  and  $R_{D^*} = 0.252 \pm 0.003$  [25].

A tree level Feynman diagram is shown in Fig. 6.1, including the SM contribution, which is mediated by a  $W^{\pm}$ , and a hypothetical charged Higgs boson contribution. Due to the difference in lepton masses, a charged Higgs contribution would be a factor  $(m_{\tau}/m_{\mu})^2 \simeq 280$  larger in the  $B \to D^{(*)}\tau^+\nu$  decay than in the  $B \to D^{(*)}\mu^+\nu$  decay, and therefore change the value of  $R_{D^{(*)}}$  [84].

Previous measurements of  $R_{D^{(*)}}$  have been made by the B-factories. The most precise measurement is from the BaBar collaboration [26], who report



Figure 6.1: Feynman diagram of  $B \to D^{(*)} \tau^+ \nu$ , including a hypothetical charged Higgs boson contribution.



Figure 6.2: Existing measurements of  $R_{D^{(*)}}$ . The SM prediction is indicated by the darker bands, the lighter band indicates the average of all measurements prior to the BaBar measurement. Figure taken from [26].

 $R_D = 0.440 \pm 0.058 \pm 0.042$  and  $R_{D^*} = 0.332 \pm 0.024 \pm 0.018$ . These results are shown in Fig. 6.2, with the SM expectations for  $R_{D^{(*)}}$  indicated by the red (dark grey) band. Previous measurements made by the Belle collaboration are also shown in Fig. 6.2, with the average indicated by the light grey band. The Belle measurements are consistent with the BaBar measurements, but have a larger uncertainty [85–87]. The Babar collaboration assess the consistency of their results with Two Higgs Doublet models (2HDM) of type II, and find that their result, together with previous bounds from  $b \rightarrow s\gamma$ , exclude the entire parameter space of this model [26]. The excess BaBar observe can be accommodated in 2HDM models of type III [88], models with new tensor mediators [89], leptoquarks [90], or coloured scalars [91]. However, the excess cannot be accommodated within the Minimal Supersymmetric Standard Model (MSSM) with MFV [92,93].

This chapter will assess the LHCb sensitivity to  $R_{D^*}$ . The same techniques could also be used to measure  $R_D$ . In order to select  $B^0 \to D^{*-}\ell^+\nu$ , the  $D^*$  meson is reconstructed via the decay chain  $D^{*+} \to D^0\pi^+$ ,  $D^0 \to K^-\pi^+$ . Unlike the other charged leptons, the  $\tau$ decays before traversing a significant portion of the detector. Due to the relatively large tau mass, the  $\tau$  can decay into final states containing hadrons, such as  $\tau^- \to \pi^-\pi^+\pi^-\nu_{\tau}$ . However, such decay modes suffer from large hadronic backgrounds and therefore in the present study only the  $\tau^- \to \mu^- \overline{\nu}_{\mu} \nu_{\tau}$  decay mode is considered. The  $B^0 \to D^{*-}\tau^+\nu$ and  $B^0 \to D^{*-}\mu^+\nu$  decays then share the same reconstructed final state; determining the  $B^0 \to D^{*-}\tau^+\nu$  yield therefore requires that the  $B^0 \to D^{*-}\mu^+\nu$  yield is also determined. The value of  $R_{D^*}$  can then be measured without relying on external input on  $\mathcal{B}(B^0 \to D^{*-}\mu^+\nu)$ , by correcting for the difference in efficiency between  $B^0 \to D^{*-}\tau^+\nu$ and  $B^0 \to D^{*-}\mu^+\nu$ .

An overview of the challenges which must be overcome in order to make a measurement of  $R_{D^*}$  is presented in Sect. 6.2. A detailed description of the backgrounds to the  $B^0 \to D^{*-} \ell^+ \nu$  decays is then presented in Sect. 6.3. The method used to determine the  $B^0 \to D^{*-} \tau^+ \nu$  signal yield is presented in Sect. 6.4, including the criteria used to select candidates. A fit to a toy dataset is then presented and this fit demonstrates the need for additional means of background rejection. Two novel algorithms to reject background decays are presented in Sect. 6.5, along with a technique to calibrate these algorithms using hadronic *B* decays selected from the data. The LHCb sensitivity to  $R_{D^*}$  is then presented in Sect. 6.6, along with an estimate of the most significant systematic uncertainties. Finally, conclusions are presented in Sect. 6.7.

# 6.2 Experimental challenges

As the neutrinos cannot be reconstructed, the  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  decays can only be selected along with all other *B* decays producing a  $D^{*-}\mu^+X$  final state, where *X* is any combination of particles. A number of *B* decays with high rates can therefore generate a background, in particular the  $B \to D^{**}\mu^+\nu$  and  $B \to D^*DX$  decays, which are described in detail in Sect. 6.3. The central challenge in measuring  $R_{D^*}$  is distinguishing the  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  decays from one another, and from this large background.

The  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  decays can be distinguished by the kinematics of the visible decay products. While neutrinos themselves are negligible in mass, the combination of the multiple neutrinos in the  $B^0 \to D^{*-}\tau^+\nu$  decays can have a large mass, in contrast to the single neutrino in  $B^0 \to D^{*-}\mu^+\nu$ . However, for other decays which produce a  $D^{*-}\mu^+X$  final state, and therefore need to be distinguished from  $B^0 \to D^{*-}\mu^+\nu$ and  $B^0 \to D^{*-}\tau^+\nu$ , the X system can also potentially have a large 'missing mass'. The missing mass is not directly measurable at a hadron collider, and so other kinematic distributions, described in Sect. 6.4, must be used to distinguish between signal and background candidates. A potential fit method to such kinematic distributions is presented in Sect. 6.4, including a fit to a toy dataset.

The  $B^0 \to D^{*-}\tau^+\nu$  decays can also be distinguished from other  $B \to D^{*-}\mu^+X$ decays by the potentially measurable distance travelled by the  $\tau$  meson before decaying ('tau flight'). Non-zero tau flight distinguishes  $B^0 \to D^{*-}\tau^+\nu$  from  $B^0 \to D^{*-}\mu^+\nu$ and  $B \to D^{**}\mu^+\nu$ . However, as charmed hadrons also travel a potentially measurable distance before decaying, the tau flight does not distinguish between  $B \to D^*DX$  and  $B^0 \to D^{*-}\tau^+\nu$ . An MVA designed to measure tau flight is described in Sect. 6.5. Many of the backgrounds to  $B^0 \to D^{*-}\tau^+\nu$  include additional particles in the final state. Isolation therefore provides a means to reject these backgrounds, and to select control samples with which to assess the modelling of such backgrounds. The isolation algorithm designed for the purpose is also described in Sect. 6.5.



Figure 6.3: Predicted spectrum of  $c\overline{u}$  bound states. Taken from Ref. [94].

# 6.3 Backgrounds

Due to the lack of a narrow peak in the reconstructed mass, the  $B^0 \to D^{*-}\tau^+\nu$  decay is sensitive to a number of classes of background events. In particular, B hadron decays of the form  $B \to D^{*-}\mu^+ X$ , referred to as 'physics backgrounds', need to be distinguished from the signal channel. The two most significant physics backgrounds are decays of the form  $B \to D^{**}\mu^+\nu$ , where  $D^{**}$  refers to any open charm meson heavier than the  $D^*$ , and decays of the form  $B \to D^{(*)}DX$ , with  $D \to \mu^+\nu X$ . These two classes of decays are described in detail in Sect. 6.3.1 and Sect. 6.3.2. Backgrounds which do not contain a genuine muon, referred to as 'misidentified muon backgrounds', are described in Sect. 6.3.3. Combinations of a  $D^*$  and a muon from different decays, referred to as 'combinatorial backgrounds', are modelled using the  $D^{*+}\mu^+$  final state, as described in Sect. 6.3.4.

# 6.3.1 $B \rightarrow D^{**} \mu^+ \nu$ decays

The modified Godfrey-Isgur predictions for the masses of  $c\overline{u}$  bound states [95] are shown in Fig. 6.3. States higher in mass than the  $D^*$  are collectively referred to as  $D^{**}$ . The measured masses and widths of the  $D^{**}$  states are summarised in Tab. 6.1.

The absolute branching fractions of  $D^{**}$  decays have not been measured. However, the products of  $\mathcal{B}(B \to D^{**}\ell^+\nu) \times \mathcal{B}(D^{**} \to D^{(*)}\pi^+)$  have been measured by several collaborations. The results of these measurements have been averaged by HFAG [6], and the results are given in Tab. 6.2. Converting these products of branching fractions into the  $B \to D^{**}\mu^+\nu$  branching fractions requires assumptions about the  $D^{**} \to D^*\pi$  branching fractions. Ref. [7] gives a detailed description of one such set of assumptions, and the resulting  $B \to D^{**}\mu^+\nu$  branching fractions are listed in Tab. 6.3. In addition, Tab. 6.3 also contains the measured  $B \to D^{(*)}\mu^+\nu$  and  $B^+ \to X_c\mu^+\nu$  branching fractions, as

State	Mass (MeV/ $c^2$ )	Width (MeV/ $c^2$ )	Reference
$D(2420)^{0}$	$2419.6 \pm 0.7$	$35.2 \pm 1.0$	Ref. [94]
$D(2460)^{0}$	$2460.4 \pm 1.3$	$43.2 \pm 3.2$	Ref. [94]
$D(2550)^{0}$	$2579.5 \pm 6.5$	$178 \pm 49$	Ref. [94]
$D(2650)^{0}$	$2649.2 \pm 4.9$	$140 \pm 25$	Ref. [94]
$D(2750)^{0}$	$2737.0 \pm 11.7$	$73 \pm 28$	Ref. [94]
$D(3000)^{0}$	$2978.1\pm8.7$	$188.1 \pm 44.8$	Ref. [94]
$D(2430)^0$	$2427 \pm 40$	$384^{+130}_{-110}$	Ref. [8]

Table 6.1: Measured masses and widths of  $D^{**}$  states with various quantum numbers.

B decay mode	Branching fraction (%)
$B^+ \to D(2420)^0 \mu^+ \nu, \ D^{**} \to D^{*-} \pi^+$	$0.29 \pm 0.02$
$B^+ \to D(2460)^0 \mu^+ \nu, \ D^{**} \to D^{*-} \pi p$	$0.07\pm0.01$
$B^+ \to D(2435)^0 \mu^+ \nu, \ D^{**} \to D^{*-} \pi^+$	$0.13 \pm 0.04$
$B^+ \to D(2400) \mu^+ \nu, \ D^{**} \to D^+ \pi$	$0.29 \pm 0.05$

Table 6.2: HFAG averages of  $\mathcal{B}(B \to D^{**}\ell^+\nu) \times \mathcal{B}(D^{**} \to D^{(*)}\pi^+)$  measurements [6].

well as a non-resonant  $D^{(*)}\pi^+$  contribution, which is denoted  $B^- \to D^{(*)}\pi^+\mu^+\nu_{NR}$ . This non-resonant contribution is calculated from the difference between the measured value of  $\mathcal{B}(B \to D^{**}\ell^+\nu) \times \mathcal{B}(D^{**} \to D^{(*)}\pi^+)$  and the inclusive  $B^- \to D^{(*)}\pi^+\mu^+\nu$  measurement [6]. Measurements of the branching fractions of the exclusive  $B \to D^{(*)}\mu^+\nu$ ,  $B \to D^{**}\mu^+\nu$  and non-resonant  $B^- \to D^{(*)}\pi^+\mu^+\nu$  final states should sum to give the measured  $B \to X_c\ell^+\nu$ branching fraction. However, a significant gap of  $1.7 \pm 0.2\%$  is seen however, indicating that some decay modes are yet to be measured [7]. Many suggestions have been made on which decays might fill this gap, with most proposing unmeasured parts of the  $B \to D^{**}\mu^+\nu$  [7,96] spectrum. As the measurement of  $B^0 \to D^{*-}\tau^+\nu$  depends upon the correct modelling of these decays, it is crucial that they be fully understood.

<i>B</i> decay mode	Branching fraction $(\%)$
$B^+ \to D^0 \mu^+ \nu$	$2.30 \pm 0.1$
$B^+ \to D^* \mu^+ \nu$	$5.34 \pm 0.12$
$B^+ \to D(2420)^0 \mu^+ \nu$	$0.65 \pm 0.07$
$B^+ \to D(2460)^0 \mu^+ \nu$	$0.28 \pm 0.03$
$B^+ \to D(2400)^0 \mu^+ \nu$	$0.44 \pm 0.08$
$B^+ \to D(2435)^0 \mu^+ \nu$	$0.20\pm0.06$
$B^+ \to D^{(*)+} \pi^+ \mu^+ \nu_{NR}$	$0.17 \pm 0.14$
$\Sigma$ (Exclusive)	$9.2 \pm 0.2$
$B^+ \to X_c \mu^+ \nu$	$10.90 \pm 0.14$
Inclusive - $\Sigma(\text{Exclusive})$	$1.7 \pm 0.24$

Table 6.3: Measured  $B^+ \to X_c \mu^+ \nu$  inclusive and exclusive decay modes, corrected for  $D^{**}$  decay modes [7].

Decay mode	Branching fraction $(\%)$
$D^+ \to \overline{K}{}^0 e^+ \nu_e$	$8.8 \pm 0.2$
$D^+ \to K^- \pi^+ e^+ \nu_e$	$4.0 \pm 0.1$
$D^+ \to \pi^0 e^+ \nu_e$	$0.41 \pm 0.02$
$D^+ \to \rho^0 e^+ \nu_e$	$0.22 \pm 0.04$
$D^+ \to \omega e^+ \nu_e$	$0.16 \pm 0.07$
$D^+ \to \eta e^+ \nu_e$	$0.11\pm0.01$
$D^+ \rightarrow e^+ X$	$16.1 \pm 0.3$

Table 6.4: Measured  $D^+ \to e^+ X$  inclusive and exclusive decay modes [8]. Assuming the  $\overline{K}{}^0\pi^0 e^+\nu_e$  final state is 50% the rate of  $K^-\pi^+e^+\nu_e$  (as is the case for the dominant  $K^*(892)^0$  resonance), the measured exclusive decay modes saturate the inclusive rate.

Decay mode	Branching fraction $(\%)$
$D^0 \to K^- e^+ \nu_e$	$3.55 \pm 0.05$
$D^0 \to K^- \pi^0 e^+ \nu_e$	$1.6^{+1.3}_{-0.5}$
$D^0 \to \overline{K}{}^0 \pi^- e^+ \nu_e$	$2.7^{+0.9}_{-0.7}$
$D^0 \to \pi^- e^+ \nu_e$	$0.29\pm 0.01$
$D^0 \to \rho e^+ \nu_e$	$0.19 \pm 0.04$
$D^0 \to e^+ \nu_e X$	$6.49 \pm 0.11$

Table 6.5: Measured  $D^0 \rightarrow e^+ X$  inclusive and exclusive decay modes [8]. The measured exclusive decay modes saturate the inclusive rate.

# 6.3.2 $B \rightarrow D^{(*)}DX$ decays

The  $B \to D^*DX$  and  $D \to \mu^+\nu X$  decays both have large branching fractions, and  $B \to D^*DX$  decays therefore constitute a significant background to the identification of  $B^0 \to D^{*-}\ell^+\nu$ . The measured  $D^+ \to e^+X$ ,  $D^0 \to e^+X$  and  $D_s^+ \to e^+X$  decay modes are listed in Tabs 6.4–6.6. The  $e^+X$  measurements are shown, rather than the  $\mu^+X$  final state, as the existing measurements are more comprehensive. In each case, the measured exclusive  $D \to e^+X$  branching fractions saturate the measured inclusive  $D \to e^+X$  rate. The  $b \to D^{*+}D^-X$ ,  $b \to D^{*+}\overline{D}{}^0X$  and  $b \to D^{*+}D_s^-X$  decays have been measured by Aleph [97], but the uncertainties are  $\mathcal{O}(30)\%$ , limiting the value of any comparison between inclusive and exclusive measurements.

Decay mode	Branching fraction $(\%)$
$D_s^+ \to \phi e^+ \nu_e$	$2.5 \pm 0.1$
$D_s^+ \to \eta^{(\prime)} e^+ \nu_e$	$3.7 \pm 0.4$
$D_s^+ \to K^0 e^+ \nu_e$	$0.4 \pm 0.1$
$D_s^+ \to K^* (892)^0 e^+ \nu_e$	$0.2 \pm 0.1$
$D_s^+ \to f_0(892)e^+\nu_e$	$0.20\pm0.03$
$D_s^+ \to e^+ \nu_e X$	$6.5 \pm 0.4$

Table 6.6: Measured  $D_s^+ \to e^+ X$  inclusive and exclusive decay modes [8]. The measured exclusive decay modes saturate the inclusive rate.



Figure 6.4: Distributions of true missing mass for simulated  $B^0 \to D^{*-}\tau^+\nu$  events, shown in black,  $B \to D^{**}\mu^+\nu$  events, shown in light grey, and  $B \to D^*DX$  events, shown in blue (darkest grey). For  $B \to D^{**}\mu^+\nu$ , the missing mass is zero.

#### 6.3.3 Misidentified muon backgrounds

The decays  $B \to D^{*+}h^{\pm}X$  generate a background to  $B^0 \to D^{*-}\ell^+\nu$  when a hadron is identified as a muon. Similarly, combinations of a  $D^*$  and a hadron originating from different decays can also generate a background. The level of these backgrounds is estimated from data using the  $D^{*-}h^{\pm}$  final state, where the hadron is required to fail the muon identification requirements described in Sect. 6.4.1. Using PID variable distributions measured from samples of  $D^{*+} \to (D^0 \to K^-\pi^+)\pi^+$  and  $\Lambda^0 \to p^+\pi^-$ , the fractions of kaons, pions and protons in the  $B \to D^{*+}h^{\pm}X$  sample can be determined. These fractions, together with the probability for a kaon, pion or proton to be misidentified as a muon, then allow the  $B \to D^{*+}h^{\pm}X$  candidates to be weighted by the probability that the hadron is misidentified as a muon, providing a measurement of the fake muon component in the  $B \to D^{*+}\mu^{\pm}$  final state [98].

#### 6.3.4 Combinatorial background

Combinatorial backgrounds are modelled using the  $D^{*+}\mu^+$  final state. The assumption that the combinatorial background in the  $D^{*+}\mu^-$  final state is well modelled by that in the  $D^{*+}\mu^+$  final state must be tested on data in any eventual measurement.

# 6.4 Signal yield determination

The kinematics of the visible particles in  $B \to D^{*-}\mu X$  decay modes depend on the mass and momentum of the missing particles. This allows  $B^0 \to D^{*-}\tau^+\nu$  to be distinguished from  $B^0 \to D^{*-}\mu^+\nu$ , and from other sources of background. The missing mass distributions

$B^0 \rightarrow D^{*-} \tau^+ \nu$	Simulated
$B^0 \rightarrow D^{*-} \mu^+ \nu$	Simulated
$B \to D(2420)\mu^+\nu$	Simulated
$B \to D(2460)\mu^+\nu$	Simulated
$B \rightarrow D(2445)\mu^+\nu$	Simulated
$B \rightarrow D^* D X$	Simulated
Combinatorial background	$D^{*+}\mu^+$ data

Table 6.7: List of PDFs included in the fit, along with the source of the events used to produce the template.

for simulated  $B^0 \to D^{*-} \tau^+ \nu$ ,  $B \to D^{**} \mu^+ \nu$  and  $B \to D^* DX$  events are shown in Fig. 6.4. Previous measurements of  $B^0 \to D^{*-} \tau^+ \nu$  have been made at electron-positron colliders, where the centre-of-mass of the  $b\bar{b}$  system is known. Reconstructing the entire event then allows the calculation of the missing mass squared. This is not possible at a hadron collider, as the centre-of-mass frames of both the B hadron and the bb system are unknown. The missing mass squared therefore cannot be used directly to distinguish between signal and background events. However, the measured kinematics of the visible particles can be used to construct variables which are correlated to the missing mass, giving sensitivity to the kinematics of the missing system. One such variable is the mass of the visible particles ('visible mass'), which has a negative correlation with the missing mass. Another variable is the corrected mass, which is defined in Eqn. 3.3, in Sect. 3.10. The corrected mass measures the minimum mass of the decaying B hadron, assuming a missing massless particle. If the unreconstructed B hadron decay products have a non-zero mass, then the assumption of a massless missing particle does not hold, and so a lower value for the corrected mass is obtained. The corrected mass therefore provides sensitivity to the missing mass. The two dimensional distributions of corrected mass and visible mass are shown in Fig. 6.5 for simulated  $B^0 \to D^{*-}\mu^+\nu$ ,  $B^0 \to D^{*-}\tau^+\nu$ ,  $B \to D^{**}\mu^+\nu$ , and  $B \to D^{**}\mu^+\nu$  $D^*DX$  events, all of which are distinguishable from one another. The two dimensional distributions of corrected mass and visible mass are therefore fitted to determine the signal yield. Histogram templates are generated from simulated events, or from control samples selected from data, and a binned extended maximum likelihood fit is performed. The events used to generate the templates are required to pass the selection requirements described in Sect. 6.4.1. The PDFs included in the fit are listed in Tab. 6.7, along with the type of sample used to generate the templates. No external measurements are used to constrain the yield of any component.

The efficiency to reconstruct and select  $B \to D^{*-}\mu^+ X$  candidates depends on the  $B \to D^{*-}\mu^+ X$  kinematics. It is therefore important that the requirements used to select candidates do not have an efficiency which differs considerably for different  $B \to D^{*-}\mu^+ X$  decay modes. The visible mass and corrected mass are fitted to determine the  $B^0 \to D^{*-}\tau^+\nu$  yield, and it is therefore crucial that the selection requirements pre-



Figure 6.5: Visible mass versus corrected mass for simulated  $B^0 \to D^{*-}\mu^+\nu$  events (top left),  $B^0 \to D^{*-}\tau^+\nu$  events (top right),  $B \to D^{**}\mu^+\nu$  events (bottom left), and  $B \to D^*DX$  events (bottom left).

serve the separation between  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  in these variables. While a variation in efficiency across the plane of visible and corrected mass could be corrected for, the separation between  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  would still be reduced. The momentum and transverse momentum of the  $D^0$  and  $\mu^+$  are shown in Fig. 6.6 and Fig. 6.7 respectively, for simulated  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  events. The muon carries a significantly lower (transverse) momentum in the  $B^0 \to D^{*-}\tau^+\nu$  decay than in the  $B \to D^{**}\mu^+\nu$  decay, indicating that the selection requirements should depend as little as possible on the muon kinematics. The differences in the  $D^0$  (transverse) momentum are less pronounced, indicating that selection criteria which depend upond the  $D^0$  kinematics do not cause a large difference in efficiency between  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$ .

#### 6.4.1 Selection criteria

The preselection criteria are listed in Tab. 6.8. These criteria are designed to minimise the efficiency difference between  $B^0 \to D^{*-} \mu^+ \nu$  and  $B^0 \to D^{*-} \tau^+ \nu$ , by avoiding requirements on biasing variables, such as the muon  $p_{\rm T}$ . After the preselection, a small number of additional requirements are applied, listed in Tab. 6.9. A requirement on the  $D^0 \chi^2_{IP}$  is applied to remove backgrounds from  $D^0$  mesons originating from the PV, reducing such backgrounds to a negligible level.

The trigger is another potential source of bias in the distributions of visible mass and



Figure 6.6: Distributions of muon momentum (left) and transverse momentum (right) for simulated  $B^0 \to D^{*-} \mu^+ \nu$  events, shown in black, and  $B^0 \to D^{*-} \tau^+ \nu$ , shown in purple (light grey).



Figure 6.7: Distributions of  $D^0$  momentum (left) and transverse momentum (right) for simulated  $B^0 \to D^{*-} \mu^+ \nu$  events, shown in black, and  $B^0 \to D^{*-} \tau^+ \nu$ , shown in purple (light grey).

Cut	Value
$\mu^+$ track $\chi^2/\text{ndof}$	< 4
$\mu^+ p$	$> 3 \mathrm{GeV}/c^2$
$\mu^+ \chi^2_{IP}$	> 45
$\mu^+$ isMuon	True
$\mu^+ \text{ DLL}_{\mu\pi}$	> 2
$K^+, \pi^- \operatorname{track} \chi^2/\operatorname{ndof}$	< 4
$K^+, \pi^- \operatorname{track} p$	$> 2 \mathrm{GeV}/c^2$
$K^+, \pi^- \operatorname{track} p_{\mathrm{T}}$	$> 0.5 \mathrm{GeV}/c^2$
$K^+, \pi^- \operatorname{track} \chi^2_{IP}$	> 45
$K^+$ DLL <sub><math>K\pi</math></sub>	> 4
$\pi^- \operatorname{DLL}_{K\pi}$	< 2
scalar sum $(K^+ p_{\rm T}, \pi^- p_{\rm T})$	> 1400  MeV/c
$D^0 \theta_P$	$< 31.6 \mathrm{mrad}$
$D^0$ vertex $\chi^2/\text{ndof}$	< 4
$D^0$ flight distance $\chi^2$	> 250
$M_{D^0}$	$1790 < m_{D^0} < 1950 \text{ MeV}/c^2$
$D^0 \mu^+ ] \theta_P$	$< 20 \mathrm{mrad}$
$D^0 \mu^+$ vertex $\chi^2$ /ndof	< 6
$M_{D^0\mu^+}$	$< 10.5 \mathrm{GeV}/c^2$

Table 6.8: The selection criteria used in the preselection.

Cut	Value
$D^0 \chi^2_{IP}$	> 10
$m_{D^0}$	$1845 < m_{D^0} < 1890 \text{ MeV}/c^2$
$m_{D^*} - m_{D^0}$	$143 < m_{D^*} - m_{D^0} < 147 \text{ MeV}/c^2$
$D^* \mu^+$ vertex $\chi^2/\text{ndof}$	< 6

Table 6.9: Selection requirements applied after the preselection.

corrected mass. These distributions are shown in Fig. 6.8 for simulated  $B^0 \to D^{*-}\mu^+\nu$ and  $B^0 \to D^{*-}\tau^+\nu$  events before and after requiring candidates to have passed topological trigger algorithms, or charm trigger algorithms, both of which are defined in Sect. 3.10. The topological trigger results in a significant shift in the visible mass and corrected mass distributions for  $B^0 \to D^{*-}\tau^+\nu$  events, without any corresponding shift for  $B^0 \to D^{*-}\mu^+\nu$ events. This is due to the topological trigger algorithm placing a requirement on the output of an MVA which gives higher output values for candidates with higher visible mass, corrected mass, and track transverse momenta. The charm trigger algorithms introduce very little change in the visible mass and corrected mass distributions for either  $B^0 \to D^{*-}\tau^+\nu$  or  $B^0 \to D^{*-}\mu^+\nu$  events. Candidates are therefore required to have passed the charm trigger algorithms. The simulation of the charm trigger algorithms can be validated using the control channels described in Sect. 6.5.3, and does not contribute to the systematic uncertainties described in Sect. 6.6.1.


Figure 6.8: Visible mass versus corrected mass for simulated  $B^0 \to D^{*-} \mu^+ \nu$  events (left column) and  $B^0 \to D^{*-} \tau^+ \nu$  events (right column), before imposing any trigger requirements (top row), after requiring candidates to have passed the topological trigger algorithms (middle row), and instead requiring candidates to have passed charm trigger algorithms (bottom row).



Figure 6.9: Visible mass (left) and corrected mass (right) projections of a toy dataset.

### 6.4.2 Fit to a toy dataset

To demonstrate the fit method, a toy dataset is generated from the samples of events listed in Tab. 6.7, after the application of the selection criteria. The expected yields for all event types are computed for the  $\sim 3 \, \text{fb}^{-1}$  of data taken in 2011 and 2012, and the corresponding number of entries are generated from the relevant template. The SM expectation for  $R_{D^*}$ is used to determine the number of  $B^0 \rightarrow D^{*-}\tau^+\nu$  decays generated. A fit to this toy sample is shown in Fig. 6.9. The  $B^0 \rightarrow D^{*-}\tau^+\nu$  component is considerably smaller than the backgrounds, and the sensitivity to  $R_{D^*}$  will therefore be poor. This motivates the use of additional methods to reduce the fraction of background decays. Such methods are presented in the next section.

## 6.5 Background rejection

In addition to the kinematic differences described in Sect. 6.4, there are two other means of distinguishing  $B^0 \to D^{*-}\tau^+\nu$  decays from background processes: the distance the tau travels before decaying, and the absence of additional charged tracks. A technique for measuring the flight of the tau is presented in Sect. 6.5.1, followed by an isolation algorithm in Sect. 6.5.2. Both of these allow the selection of a sample with an increased fraction of  $B^0 \to D^{*-}\tau^+\nu$  decays relative to other processes. Additionally, both techniques also allow samples to be selected which have enhanced fractions of background processes. This allows the modelling of such backgrounds to be validated.

### 6.5.1 Displaced $\tau$ decays

One characteristic differentiating  $B^0 \to D^{*-}\tau^+\nu$  decays from  $B^0 \to D^{*-}\mu^+\nu$  and  $B \to D^{**}\mu^+\nu$  decays is the lifetime of the  $\tau$ . However, the  $\tau$  decay vertex position cannot be measured in  $\tau^+ \to \mu^+\nu\nu$ , as there is only one charged track. The only case where the flight of the  $\tau$  can be identified is where it has travelled a greater distance than the  $D^0$ . This occurs in  $\sim 25\%$  of  $B^0 \to D^{*-}\tau^+\nu$  decays. In this case the  $[D^0\mu]$  vertex may be



Figure 6.10: Schematic diagram of the decays  $B^0 \to D^{*-}\mu^+\nu$  (top), and  $B^0 \to D^{*-}\tau^+\nu$ , where the  $\tau$  travels a greater distance than the  $D^0$  before decaying (bottom). The  $D^0$  and  $\mu$  trajectories are indicated by solid arrows, with the solid ellipse corresponding to the  $[D^0\mu]$  vertex position which in the  $B^0 \to D^{*-}\tau^+\nu$  case is downstream of the  $[D^0]$  decay position.



Figure 6.11: Difference in the true decay positions between the  $\tau$  and the  $D^0$  for simulated  $B^0 \to D^{*-}\mu^+\nu$ , shown in black, and  $B^0 \to D^{*-}\tau^+\nu$ , shown in purple (light grey) events. Distances are shown in the direction parallel (left) and perpendicular (right) to the beam direction.



Figure 6.12: Difference in the measured decay positions between the  $\tau$  and the  $D^0$  for simulated  $B^0 \to D^{*-}\mu^+\nu$ , shown in black, and  $B^0 \to D^{*-}\tau^+\nu$ , shown in purple (light grey) events. Distances are shown in the direction parallel (left) and perpendicular (right) to the beam direction.

located downstream of the  $[D^0]$  vertex, as illustrated in Fig. 6.10. This corresponds to measuring a negative  $D^0$  lifetime, and so is never physical for  $B \to D\mu^+ X$  modes. The difference between the true decay positions of the  $\tau$  and the  $D^0$  is shown in Fig. 6.11 for simulated  $B^0 \to D^{*-}\tau^+\nu$  and  $B^0 \to D^{*-}\mu^+\nu$  events, before the application of any selection criteria. As shown in Fig. 6.12, resolution effects result in a small number of  $B^0 \to D^{*-}\mu^+\nu$  candidates with measured negative  $D^0$  lifetimes.

The distributions of  $D^0\mu^+$  vertex  $\chi^2$  is shown for simulated  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  events in Fig. 6.13. The requirement on this quantity imposed in the preselection is indicated by the vertical line in the figure. The  $\tau$  flight does not result in a significantly larger fraction of  $B^0 \to D^{*-}\tau^+\nu$  decays relative to  $B^0 \to D^{*-}\mu^+\nu$  decays being removed by the vertex quality requirement.

A BDT is trained to use the  $\tau$  decay length to distinguish between samples of simulated  $B^0 \to D^{*-} \tau^+ \nu$  and  $B^0 \to D^{*-} \mu^+ \nu$  events. The input variables are listed in Tab. 6.10. Variables with a significant correlation to the visible mass are excluded, in order to ensure



Figure 6.13: Distribution of  $D^0\mu$  vertex  $\chi^2$  for  $B^0 \to D^{*-}\mu^+\nu$ , shown in black, and  $B^0 \to D^{*-}\tau^+\nu$ , shown in light grey.



Table 6.10: Input variables for the flight BDT.

the correlation with the variables used to separate  $B^0 \to D^{*-} \tau^+ \nu$  from  $B^0 \to D^{*-} \mu^+ \nu$  in Sect. 6.6 is minimal.

The BDT output distribution for simulated  $B^0 \to D^{*-}\mu^+\nu$  and  $B^0 \to D^{*-}\tau^+\nu$  events is shown in Fig. 6.14. For a  $B^0 \to D^{*-}\tau^+\nu$  efficiency of 20% (5%), the  $B^0 \to D^{*-}\mu^+\nu$ efficiency is 4% (0.2%). This demonstrates that the tau flight has considerable power in rejecting  $B^0 \to D^{*-}\mu^+\nu$  backgrounds, which has not previously been used in any  $B \to \tau$ analyses, at LHCb or elsewhere. The MVA approach used here rejects roughly a factor two more  $B^0 \to D^{*-}\mu^+\nu$  for a given  $B^0 \to D^{*-}\tau^+\nu$  efficiency compared to placing a requirement on the distance between the  $[D^0\mu]$  and  $[D^0]$  vertices.

#### 6.5.2 Multivariate isolation

Many background decays contain additional reconstructible particles which can be used to veto such decays. Only additional charged particles are considered, due the greater precision with which such particles are measured at LHCb.

Previous isolation methods used in heavy flavour measurements have fallen into two categories: those based around summing properties of the tracks within an angular cone



Figure 6.14: Distribution of flight MVA output for simulated  $B^0 \to D^{*-} \mu^+ \nu$  (red hatched) and  $B^0 \to D^{*-} \mu^+ \nu$  (blue) events.

around the *B* candidate, or around the final state particles (cone isolation) [99–101]; and those based on the proximity of the closest track to the *B* candidate, as defined by the vertex  $\chi^2$  (vertex isolation), or the DOCA [101]. The cone isolation approach has the disadvantage that, in general, it includes a large number of tracks, and so is not very sensitive to backgrounds with one or two tracks missing. The vertex isolation approach allows individual tracks to be identified as originating from the same *B* meson as the signal *B* candidate, but is sensitive to tracks originating from the PV unless additional requirements are imposed.

A new approach is introduced here, employing an MVA to determine whether a given track originates from the same B meson as the B candidate (referred to as 'associated tracks'), or from anywhere else from the rest of the event (referred to as 'unassociated tracks'). Input variables to the MVA are listed in Tab. 6.11, and include the properties of the track, and the properties of the candidate B decay with the vertex refitted to include the track that is considered. Long tracks, VELO tracks and upstream tracks (defined in Sect. 3.6) are all included, in order to find the maximum possible number of associated tracks. This MVA is applied to each track in the event, and the tracks with the highest (most associated-track like) MVA output are considered for further analysis.

The MVA is trained using associated tracks taken from the  $D^{**} \to D^*$  decay in simulated  $B \to D^{**}\mu^+\nu$  events, and using unassociated tracks taken from simulated  $B^0 \to D^{*-}\mu^+\nu$  events with the signal decay excluded. In both cases, *B* candidates are reconstructed in the  $D^*\mu$  final state. The output distribution for this MVA is shown in Fig. 6.15 for simulated  $B \to D^{**}\mu^+\nu$  and  $B^0 \to D^{*-}\mu^+\nu$  events. The presence of  $D^{**} \to D^*$ decay modes with no charged tracks results in a component of the  $B \to D^{**}\mu^+\nu$  MVA



Figure 6.15: Highest isolation MVA output for any track in simulated  $B^0 \to D^{*-} \mu^+ \nu$ events (magenta) and simulated  $B \to D^{**} \mu^+ \nu$  events (red hatched).

output distribution matching that of  $B^0 \to D^{*-}\mu^+\nu$ , which peaks at lower MVA output values. The performance of placing a requirement on the highest MVA track is compared to the standard LHCb vertex isolation tool in Fig. 6.16. The gain in background rejection around a factor 2 (1.2) at a signal efficiency of 95% (75%).

In addition to rejecting backgrounds, the isolation MVA can be used as a selection tool. For example, the MVA output can be used to add pions to  $D^{*-}\mu^+$  candidates in data, in order to search for  $B \to D^{**}\mu^+\nu$  decays. The  $D^{*-}\mu^+$  candidates are taken from ~ 3 fb<sup>-1</sup> of data taken in 2011 and 2012, using the selection criteria listed in Sect. 6.4.1. No momentum information is available for VELO tracks, and such tracks are therefore excluded. No PID or charge requirements are placed on the pion. The resulting  $M_{D^{*-}\pi^+}$  distributions are shown in Fig. 6.17 for the two highest MVA output tracks in the event, with a minimum MVA output requirement of 0.3 imposed. For the highest MVA output track, a peak is visible around  $M_{D^{*-}\pi^+} \sim 2400$  from the  $D(2420) \to D^{*-}\pi^+$  and  $D(2460) \to D^{*-}\pi^+$ decays. This peak is vastly diminished for the second highest MVA output track, showing that the number of cases where the isolation MVA selects an unassociated track before a associated track is low. This demonstrates that the isolation MVA has the power to distinguish between partially reconstructed signal *B* decays and unrelated tracks.

## 6.5.3 Hadronic control channels

Fully reconstructed B decays such as  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  can be used as control channels for  $B^0 \to D^{*-}\tau^+\nu$  and  $B^0 \to D^{*-}\mu^+\nu$ , in order to calibrate the output of the flight BDT and isolation BDT. The  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  candidates are selected from ~ 3 fb<sup>-1</sup> of data taken in 2011 and 2012, and are then separated from backgrounds using a fit to the  $M_{D^{*-}\pi^+\pi^-\pi^+}$  distribution. The signal is modelled using a Crystal Ball PDF, with



Figure 6.16: Comparison in performance between vertex isolation, shown in cyan (dark grey) and the novel MVA isolation, shown in blue (light grey).

Track properties:
$IP\chi^2_{PV}$
$IP\chi^2_{SV}$
$p_{\mathrm{T}}$
$\cos(D^{(*)}\mu,\mathrm{track})$
Vertex properties:
$FD\chi^2$
$\Delta FD\chi^2$

Table 6.11: Input variables for the isolation MVA



Figure 6.17: Distributions of  $M_{D^{*-}\pi^+}$ , where the isolation MVA is used to add a pion to  $D^{*-}\mu^+$  candidates in data. Distributions are shown for track in the event with the highest MVA output (top), and the next highest MVA output (bottom). A peak from  $D(2420) \rightarrow D^{*-}\pi^+$  and  $D(2460) \rightarrow D^{*-}\pi^+$  decays is visible for the candidate with the highest MVA output, but is considerably smaller for the candidate with the second highest output.



Figure 6.18: The  $M_{D^{*-}\pi^+\pi^-\pi^+}$  distribution of  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  candidates. The fit model is described in the text.

an additional Gaussian component sharing the same mean, but differing in width. The combinatorial background is modelled using an exponential PDF. The results of the fit are shown in Fig. 6.18. After the fit, the sPlot technique [102] is used to produce background subtracted distributions.

The isolation BDT output distribution is shown in Fig. 6.19 for  $B^0 \rightarrow D^{*-}\pi^+\pi^-\pi^+$ candidates taken from background subtracted data, and from simulated events. Reasonable agreement between data and simulation is seen, without any correction applied to the simulated events.

By vertexing a subset of the final state particles,  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  can be made to more closely resemble the topology of  $B^0 \to D^{*-}\tau^+\nu$  or  $B^0 \to D^{*-}\mu^+\nu$ , and so can be used to calibrate the flight MVA output for these decay modes. In particular, by constructing a vertex consisting of the three pions produced directly in the *B* decay (referred to as the  $H^+$ ) and the kaon from the  $D^0$  decay (referred to as the  $[K^-H^+]$  vertex), the topology of  $B^0 \to D^{*-}\tau^+\nu$  can be mimicked. This arrangement is illustrated in Fig. 6.20, and allows the performance of the flight BDT on  $B^0 \to D^{*-}\tau^+\nu$  to be calibrated from the data. In addition, by fitting the vertex consisting of the  $D^*$  and one of the pions (referred to as the  $[D^{*-}\pi^+]$  vertex), the flight BDT performance on  $B^0 \to D^{*-}\mu^+\nu$  can be calibrated. The background subtracted flight BDT output distributions are shown for these two configurations in Fig. 6.21, for  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  candidates taken from backgroundsubtracted data, and from simulated events. Again, reasonable agreement between data and simulation is seen, without any correction applied to the simulated events.



Figure 6.19: Isolation BDT output distributions for  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  candidates taken from background-subtracted data, shown in black, and simulated events, shown in purple (light grey).



Figure 6.20: Schematic diagrams of the  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  decay, reconstructed as  $[D^{*-}\pi^+]$  (top), and  $[K^-H^+]$  (bottom). The  $[K^-H^+]$  vertex is reconstructed downstream of the  $[H^+]$  vertex, mimicking the topology of  $B^0 \to D^{*-}\tau^+\nu$ .



Figure 6.21: Flight BDT output distributions for  $B^0 \to D^{*-}\pi^+\pi^-\pi^+$  candidates taken from background-subtracted data, shown in black, and simulated events, shown in purple (light grey). The flight BDT is evaluated using the  $[D^{*-}\pi^+]$  vertex, mimicking  $B^0 \to D^{*-}\mu^+\nu$ (left), and the  $[K^-H^+]$  vertex, mimicking  $B^0 \to D^{*-}\tau^+\nu$  (right).

## 6.6 Sensitivity to $R_{D^*}$

The two dimensional distribution of visible mass and corrected mass offers a means of separating  $B^0 \to D^{*-} \tau^+ \nu$  decays from  $B^0 \to D^{*-} \mu^+ \nu$  decays, and from other sources of background events. A template fit to these distributions is used to extract the  $B^0 \rightarrow D^{*-} \tau^+ \nu$ and  $B^0 \to D^{*-} \mu^+ \nu$  yields, as described in Sect. 6.4. The isolation BDT and flight BDT described in Sects. 6.5.1 and 6.5.2 respectively provide a means of selecting samples with an increased signal fraction. In addition, regions with an increased fraction of particular backgrounds can also be selected, in order to verify that all of the backgrounds are modelled correctly. This is achieved by performing the fit simultaneously in different regions of flight and isolation BDT output. A binning scheme in flight and isolation BDT output is given in Tab. 6.12. Bins are referred to by a number and a letter, referring to the region of flight BDT output and isolation BDT output respectively. The region of flight BDT is denoted by a number between zero and two, with zero referring to candidates in which no  $\tau$  flight is measured, and two referring to candidates with a large measured  $\tau$  flight. The region of isolation is denoted by either 'I' or 'A', referring to isolated and anti-isolated regions, respectively. Fits to toy datasets in each region of BDT output are shown in Figs. 6.22-6.27. With increasing flight BDT output, the fraction of  $B^0 \to D^{*-} \tau^+ \nu$  increases, but the fraction of  $B \to D^* DX$  increases faster. This is demonstrated in Fig. 6.22, 6.24, and 6.26, showing regions 0I, 1I and 2I respectively. Region 0I contains the majority of the candidates in the data sample, without any enhancement in the  $B^0 \to D^{*-} \tau^+ \nu$  fraction. In region 1I, the  $B^0 \to D^{*-} \tau^+ \nu$  signal fraction is considerably increased, and the  $B \to D^*DX$  constitutes a significant background but these events are not yet dominant. By region 2I, the  $B \to D^*DX$  dominates the sample, with  $B^0 \to D^{*-} \mu^+ \nu$  and  $B \to D^{**} \mu^+ \nu$  decays strongly suppressed relative to  $B^0 \to D^{*-} \tau^+ \nu$ decays.

The anti-isolated regions (Fig. 6.23, 6.25, and 6.27, showing regions 0A, 1A and 2A respectively) have considerably higher fractions of  $B \to D^{**}\mu^+\nu$  and  $B \to D^*DX$  relative to the modes without additional charged particles (e.g.  $B^0 \to D^{*-}\tau^+\nu$  and  $B^0 \to D^{*-}\mu^+\nu$ ). These regions therefore provide control samples in which the quality of the modelling of the  $B \to D^{**}\mu^+\nu$  and  $B \to D^*DX$  decay modes can be determined. With large measured  $\tau$  flight (region 2A, Fig. 6.27), the  $B \to D^{**}\mu^+\nu$  is strongly suppressed relative to  $B \to D^*DX$ , enabling the modelling of the  $B \to D^{**}\mu^+\nu$  and  $B \to D^*DX$  backgrounds to be assessed separately to  $B \to D^{**}\mu^+\nu$ .

The ratio  $R_{D^*}$  can be measured directly from a simultaneous fit to bins 0I, 1I and 2I, by correcting the  $B^0 \to D^{*-} \tau^+ \nu$  yield by the efficiency for it to fall in a given bin. The  $B^0 \to D^{*-} \mu^+ \nu$  yield is measured solely from bin 0I, where it is the dominant decay. Using this toy data, the statistical uncertainty on  $R_{D^*}$  is estimated to be ~ 6%, competitive with existing measurements. The sources of systematic uncertainty are considered in the

Name	Flight BDT	Isolation BDT	Toy data
IO	< 0.3	< 0.3	Fig. 6.22
0A	< 0.3	> 0.3	Fig. 6.23
1I	0.3 < BDT < 0.8	< 0	Fig. 6.24
1A	0.3 < BDT < 0.8	> 0	Fig. 6.25
2I	>0.8	< 0	Fig. 6.26
2A	>0.8	> 0	Fig. 6.27

Table 6.12: Definition of the binning scheme in flight BDT output and isolation BDT output.



Figure 6.22: Visible mass (left) and corrected mass (right) projections of a toy dataset in region 0I. This region requires candidates to be isolated, and have no measured tau flight.

next section.

#### 6.6.1 Systematic uncertainties

The largest source of systematic uncertainty is the modelling of the various background components. As described in Sect. 6.3.1, there is a significant gap between the measured exclusive and inclusive  $B \to X_c \ell^+ \nu$  decays. These missing  $D^{**}$  states which fill this gap must be measured in data. In Sect. 6.5.2, the use of the isolation BDT to add pions to  $D^{*-}\mu^+$  candidates was demonstrated. Tighter selection criteria, listed in Tabs. 6.13 and 6.14, are applied to select the  $D^{*-}\pi^+$  and  $D^{*-}\pi^+\pi^-$  final states. The resulting  $M_{D^{*-}\pi^+}$  and  $M_{D^{*-}\pi^+\pi^-}$  distributions are shown in Figs. 6.28 and 6.29. Measuring the  $D^{**}$  states in these mass spectra, together with those in the  $D^0\pi^+(\pi^-)$  spectra, should give insight into the inclusive-exclusive gap. These four mass spectra will dictate which  $B \to D^{**}\mu^+\nu$  decays need to be modelled in order to measure  $R_{D^*}$ . Simulated samples for any  $B \to D^{**}\mu^+\nu$ decays which have not previously been measured will then need to be produced.

The form factors for  $B^0 \to D^{*-}\mu^+\nu$  have been precisely measured, and so the modelling of  $B^0 \to D^{*-}\mu^+\nu$  will not give rise to a significant systematic uncertainty. The form factors for  $B \to D^{**}\mu^+\nu$  decays, however, have not been measured, and only theoretical predictions are available. It is therefore important to consider the uncertainty from the limited knowledge of the  $B \to D^{**}\mu^+\nu$  form factors. The simulated  $B \to D^{**}\mu^+\nu$  events



Figure 6.23: Visible mass (left) and corrected mass (right) projections of a toy dataset in region 0A. This region requires candidates to be anti-isolated, and have no measured tau flight.



Figure 6.24: Visible mass (left) and corrected mass (right) projections of a toy dataset in region 1I. This region requires candidates to be isolated, and have moderate measured tau flight.



Figure 6.25: Visible mass (left) and corrected mass (right) projections of a toy dataset in region 1A. This region requires candidates to be anti-isolated, and have moderate measured tau flight.

Cut	Value
$\pi^+$ isolation BDT	> 0.5
next highest isolation BDT	< 0.3
$\pi \text{ DLL}_{K\pi}$	< 0

Table 6.13: Additional requirements applied to select the  $D^{*-}\pi^+\mu^+\nu$  final state.



Figure 6.26: Visible mass (left) and corrected mass (right) projections of a toy dataset in region 2I. This region requires candidates to be isolated, and have high measured tau flight.



Figure 6.27: Visible mass (left) and corrected mass (right) projections of a toy dataset in region 2A. This region requires candidates to be anti-isolated, and have high measured tau flight.

Cut	Value
$\pi$ max isolation BDT	> 0.3
$\pi$ min isolation BDT	> 0.2
next highest isolation BDT	< 0.2
$\pi \text{ DLL}_{K\pi}$	< 0

Table 6.14: Additional requirements applied to select the  $D^{*-}\pi^+\pi^-\mu^+\nu$  final state.



Figure 6.28: Distribution of  $M_{D^{*-}\pi^+}$  for  $D^{*+}\pi^-\mu^-\nu$  candidates.



Figure 6.29: Distribution of  $M_{D^{*-}\pi^+\pi^-}$  for  $D^{*-}\pi^+\pi^-\mu^+\nu$  candidates.

are generated using the ISGW2 model [103]. Variations to this model are parameterised by coefficients multiplying the four ISGW2 form factors. These coefficients do not have any kinematic dependence.

To assess the uncertainty from the  $B \to D^{**} \mu^+ \nu$  form factors, the fit is repeated with the template for  $B \to D^{**} \mu^+ \nu$  decays modified to match different values for the relevant form factors. This modification in shape is performed by reweighting the sample of simulated  $B \to D^{**} \mu^+ \nu$  events to match other samples generated with different values of the  $B \to D^{**} \mu^+ \nu$  form factors. These additional simulated samples are generated without emulating the detector response. The values of the form factors used to generate these samples are independently varied between 0.5 and 1.5. The anti-isolation samples described in Sect. 6.6 have an increased fraction of  $B \to D^{**} \mu^+ \nu$ , and large variations of the  $B^0 \to D^{*-} \mu^+ \nu$  form factors will therefore give a poor fit to these samples, and do not need to be considered. The largest variation in the fitted value of  $R_{D^*}$  for any set of form factor values giving a fit likelihood that is within  $2\sigma$  of the lowest likelihood for any sample is taken as an estimate of the systematic uncertainty. The distribution of  $R_{D^*}$ values for form factor variations giving a fit likelihood within  $2\sigma$  of the lowest likelihood is shown in Fig. 6.30. Taking the largest variation gives a 2.9% systematic uncertainty on  $R_{D^*}$ .

The distribution of  $R_{D^*}$  values for form factor variations giving a fit likelihood within  $2\sigma$  of the lowest likelihood is shown in Fig. 6.30. Taking the largest variation gives a 2.9% systematic uncertainty on  $R_{D^*}$ .

For simplicity, this study models the entire measured  $B \to D^{**}\mu^+\nu$  rate using only the  $B \to D(2460)\mu^+\nu$  decay. The missing component of  $B \to X_c\mu\nu$  has an unknown distribution between the  $D\mu^+X$  and  $D^*\mu^+X$  final states. The missing component is assumed to be of the same rate as the measured component, and will therefore give rise to a systematic uncertainty of the same size. Similarly, the rate of  $B \to D^*DX$  background decays is slightly smaller than the rate of the measured  $B \to D^{**}\mu^+\nu$  component, and therefore the systematic uncertainty arising from the modelling of  $B \to D^*DX$  decays may also be assumed to be of a similar size to the measured  $B \to D^{**}\mu^+\nu$  component. Assuming that these three contributions may be added in quadrature, the total systematic uncertainty from the modelling of these physics backgrounds might then be expected to be ~ 5%, which is comparable to the expected statistical uncertainty on  $R_{D^*}$ .

The combinatorial background and fake muon background are both considerably smaller than the physics backgrounds described above, and so should result in lower systematic uncertainties. The combinatorial background is assumed to be well modelled by the  $D^{*+}\mu^+$  final state; if this assumption does not hold when checked in data then a systematic uncertainty will need to be assigned. In the region of  $M_{D^{*-}\mu^+}$  above the *B* mass the combinatorial background appears to be identical in yield for the  $D^{*+}\mu^+$  and



Figure 6.30: Distribution of  $R_{D^*}$  measured from a toy dataset fitted with different values used for the  $B \to D^{**} \mu^+ \nu$  form factors. Only points with a fit likelihood within  $2\sigma$  of the minimum are shown.

 $D^{*+}\mu^{-}$  final states, with visible and corrected mass distributions in good agreement, and no systematic is therefore currently assigned. However, this will require confirmation in any future measurement.

The modelling of the flight and isolation MVA output distributions will also form sources of systematic uncertainty. This uncertainty can be estimated using hadronic control channels in data, as described in Sect. 6.5.3. Taking the isolation MVA output distributions from  $B^0 \rightarrow D^{*-}\pi^+\pi^-\pi^+$  decays selected from data and from simulated events, the ratio of efficiencies for an isolation MVA requirement of > 0 is  $1.00 \pm 0.04$  between data and simulation. Similarly, the ratio of efficiencies of a flight MVA requirement of > 0 is  $0.96 \pm 0.05$ . Both of these values agree within a statistical uncertainty of ~ 5%. Assessing the systematic uncertainty, which should not be larger than ~ 5%, will require larger samples. If a disagreement close to the current statistical uncertainty is seen, then corrections will be applied to the simulated results, which should bring the resulting systematic uncertainty below that arising from the background modelling. The systematic uncertainties from sources such as the trigger, PID efficiencies and track reconstruction efficiencies may be assessed using fully reconstructed control channels and, as with the  $B^+ \rightarrow h^- \mu^+ \mu^+$  and  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  decays in Chapters 4 and 5, the resulting uncertainties will be small.

## 6.7 Conclusion

Measuring the  $B^0 \to D^{*-} \tau^+ \nu$  decay is challenging at a hadron collider, as the backgrounds are large and, in many cases, poorly measured. The  $\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$  mode allows the direct measurement of  $R_{D^*}$ . The method presented here would allow the  $B^0 \to D^{*-} \tau^+ \nu$  yield to be determined from a fit to reconstructed kinematic distributions. Additionally, novel MVA methods allow background decays to be rejected by requiring candidates are well isolated, and for  $B^0 \to D^{*-} \tau^+ \nu$  decays to be identified using the distance the  $\tau$  travels before decaying. Both the isolation and tau flight tools allow samples to be selected with a much higher fraction of  $B^0 \to D^{*-} \tau^+ \nu$  decays. In addition, these methods allow samples to be selected which have much higher fractions of background events, allowing the modelling of these backgrounds to be tested. In particular, the isolation can be used as a tool to select  $B \to D^{**} \mu^+ \nu$  decays, allowing measurements beyond those made by the B factories. The enhanced background samples introduced in this chapter will allow the uncertainty arising due to the background modelling to be estimated using data-driven methods. The estimated uncertainty on  $R_{D^*}$ , including the largest systematic uncertainties, is ~ 8%, competitive with the 9% uncertainty on  $R_{D^*}$  measured by the Babar collaboration. Further studies may be able to reduce the  $\sim 5\%$  systematic uncertainty. The study presented in this chapter is focused on  $R_{D^*}$ , but the same measurement techniques could also be applied to measure  $R_D$ . Performing a simultaneous measurement using the  $D^0 \mu^+ \nu$  final state, which contains a feed-down component from  $B^0 \to D^{*-} \ell^+ \nu$ , will improve the statistical uncertainty on  $R_{D^*}$ . Additionally, many of the same backgrounds are shared between the  $D^0\mu^+\nu$  and  $D^*\mu^+\nu$  final states, and the systematic uncertainties arising from the background modelling will therefore be reduced. In future, larger data samples will be produced, which will reduce the statistical uncertainty on  $R_{D^{(*)}}$ . The largest systematic uncertainties, from the modelling of the backgrounds and the isolation and flight MVAs, are estimated using control samples in data, and may therefore also be reduced with larger data samples. Determining to what level the systematic uncertainties can eventually be reduced will require a detailed study of the relevant control samples in data.

# Chapter 7

## Conclusions

This thesis presents two separate searches for new physics at LHCb, and a sensitivity study for a third new physics search. The limits on the  $B^+ \rightarrow h^- \mu^+ \mu^+$  branching fractions detailed in Chapter 4 are an order of magnitude improvement over previous searches; however, these limits remain at least five orders of magnitude above the region motivated by the  $\nu MSM$  model described in Chapter 2. Recent astrophysical x-ray measurements hint at a 3.5 keV x-ray spectral line which could correspond to a 7 keV sterile neutrino [104, 105], consistent with the dark matter candidate proposed by the  $\nu MSM$ . This reinforces the motivation to probe the theoretically favoured region of the  $\nu MSM$ , and a dedicated fixed-target experiment has recently been proposed, operating in the neutrino mass range accessible with D meson decays [64].

The first observation of the  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  decay was presented in Chapter 5, and the branching fraction measured. This is the first time any  $b \rightarrow d\mu^+\mu^-$  transition has been observed. A novel determination of  $|V_{\rm td}|/|V_{\rm ts}|$  was made by measuring the ratio of  $B^+ \to \pi^+ \mu^+ \mu^-$  and  $B^+ \to K^+ \mu^+ \mu^-$  branching fractions. Both of these measurements are consistent with the SM expectation, within the large statistical uncertainties. Since the publication of this measurement, hints of new physics in  $b \to s \mu^+ \mu^-$  decays have emerged, via a measurement of the angular observable  $P'_5$  [106]. This observable relates to the interference between the longitudinally and transversely polarised amplitudes for  $B^0 \to K^{*0} \mu^+ \mu^-$  [107]. The measurements of  $P'_5$  are shown in Fig. 7.1 as a function of dimuon mass, with the SM predictions taken from Ref. [108] indicated by the solid rectangles. However, other theory predictions give an uncertainty on the SM expectation for  $P'_5$  which is larger than that shown in the figure [109]. As a result of this, the significance of the discrepancy from the SM in global fits of  $b \rightarrow s$  observables varies between  $4\sigma$  [110, 111] and  $2\sigma$  [112]. A proposed explanation for the discrepancy in  $P'_5$  is a new physics contribution to the Wilson Coefficient  $C_9$ . This contribution cannot be accomodated in the MSSM, or models with partial compositeness, but can arise in models with a new heavy vector particle [110]. Independent of these global fits, a discrepancy has also emerged between measurements of the branching fractions of exclusive  $b \rightarrow s\mu^+\mu^-$  decays



Figure 7.1: Measured value of the  $B^0 \to K^{*0} \mu^+ \mu^-$  angular observable  $P'_5$  as a function of dimuon mass. The solid boxes indicate the SM expectation.

and new precise predictions at high dimuon mass made using lattice QCD [113,114]. The disagreement between the branching fractions and the lattice predictions is illustrated in Fig. 7.2. The red dashed line in Fig. 7.2 indicates the expectation for the value of  $C_9$  favoured by the  $P'_5$  measurements. The measurements are in much better agreement with the new physics prediction than the SM expectation, providing an independent hint for new physics in  $b \to s\mu^+\mu^-$ . If the equivalent discrepancy were to be observed in future measurements of  $\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-)$  this would support the tentative indication of new physics, and indicate that such new physics is minimally flavour violating. Conversely, if new physics in  $b \to s\mu^+\mu^-$  is confirmed and the  $B^+ \to \pi^+\mu^+\mu^-$  branching fraction is measured to be agreement with, or in excess of, the SM expectation, it would indicate that the new physics is non-minimally flavour violating. In either case, an update of the  $B^+ \to \pi^+\mu^+\mu^-$  measurement presented in this thesis will be of interest in the future.

In Chapter 6 a sensitivity study for a measurement of  $\mathcal{B}(B^0 \to D^{*-}\tau^+\nu)$  was presented. Such a measurement would be the first measurement involving a  $B \to \tau^-$  decay at a hadron collider. The novel fit method, isolation algorithm and tau flight algorithm presented make this measurement feasible. The isolation and flight algorithms also allow samples to be selected containing increased fractions of particular background types, allowing a data driven test of the modelling of these backgrounds. In particular, the isolation algorithm will allow the  $B \to D^{**}\mu^+\nu$  decay modes to be measured, which should give insight into the long-standing gap between inclusive and exclusive  $B \to X_c \mu^+ \nu$  decays. The study presented in this thesis focuses solely on  $B^0 \to D^{*-}\tau^+\nu$ , however the same techniques can also be used to measure  $B \to D^0 \tau^+ \nu$ . Current measurements of the ratio of  $B \to D^{(*)} \tau^+ \nu$  and  $B \to D^{(*)} \mu^+ \nu$  branching fractions remain  $3\sigma$  above the SM expectation, providing a strong motivation to measure this quantity at LHCb. Confirming



Figure 7.2: Ratio between experimental measurement and lattice QCD SM expectation for exclusive  $b \rightarrow s\mu^+\mu^-$  decays at high dimuon mass. The dashed red line indicates a new physics prediction taken from a global fit to other  $b \rightarrow s\mu^+\mu^-$  observables. Figure taken from Ref. [115].

this discrepancy would indicate the presence of new physics. As with the discrepancy in  $b \rightarrow s\mu^+\mu^-$ , the excess in  $B \rightarrow D^{(*)}\tau^+\nu$  cannot be accomodated in the MSSM [92,93].

# Bibliography

- [1] LHCb Collaboration, Search for the lepton number violating decays  $B^+ \to \pi^- \mu^+ \mu^+$ and  $B^+ \to K^- \mu^+ \mu^+$ , Phys. Rev. Lett. **108** (2011) 101601. 7 p.
- [2] LHCb Collaboration, First observation of the decay  $B^+ \to \pi^+ \mu^+ \mu^-$ , J. High Energy Phys. **12** (2012) 125. 20 p.
- [3] M. Shaposhnikov, Sterile neutrinos in cosmology and how to find them in the lab,
  J. Phys. Conf. Ser. 136 (2008) 022045, arXiv:0809.2028.
- [4] G. Cvetic, C. Dib, S. K. Kang, and C. Kim, Probing Majorana neutrinos in rare K and D, D<sub>s</sub>, B, B<sub>c</sub> meson decays, Phys. Rev. D82 (2010) 053010, arXiv:1005.4282.
- [5] G. Isidori, Y. Nir and G. Perez, Flavor physics constraints for physics beyond the Standard Model, Ann. Rev. Nucl. Part. Sci. 60 (2010) 355, arXiv:1002.0900.
- [6] Heavy Flavor Averaging Group, Y. Amhis et al., Averages of B-Hadron, C-Hadron, and tau-lepton properties as of early 2012, arXiv:1207.1158.
- [7] F. U. Bernlochner, D. Biedermann, H. Lacker, and T. Lück, Constraints on exclusive branching fractions BF<sub>i</sub>(B<sup>+</sup> → X<sup>i</sup><sub>c</sub>l<sup>+</sup>ν) from moment measurements in inclusive B → X<sub>c</sub>lν decays, arXiv:1402.2849.
- [8] Particle Data Group, J. Beringer et al., Review of particle physics, Phys. Rev. D86 (2012) 010001.
- [9] C. Burgess and G. Moore, *The Standard Model: A Primer*, Cambridge University Press, 2006.
- [10] R. K. Ellis, W. J. Stirling, and B. R. Webber, QCD and Collider Physics, Cambridge University Press, 1996.
- [11] M. Kobayashi and T. Maskawa, CP-violation in the renormalizable theory of weak interaction, Progress of Theoretical Physics 49 (1973) 652.
- [12] T. Asaka, S. Blanchet, and M. Shaposhnikov, The νMSM, dark matter and neutrino masses, Phys. Lett. B631 (2005) 151, arXiv:hep-ph/0503065.

- [13] M. Shaposhnikov, Is there a new physics between electroweak and Planck scales?, arXiv:0708.3550.
- [14] A. Boyarsky, O. Ruchayskiy, and M. Shaposhnikov, The role of sterile neutrinos in cosmology and astrophysics, Ann. Rev. Nucl. Part. Sci. 59 (2009) 191, arXiv:0901.0011.
- [15] F. L. Bezrukov, νMSM and its experimental tests, J. Phys. Conf. Ser. 110 (2008) 082002, arXiv:0710.2501.
- [16] E. Fermi, An attempt of a theory of beta radiation. 1, Z. Phys. 88 (1934) 161.
- [17] W. Altmannshofer et al., Symmetries and asymmetries of  $B \to K^* \mu^+ \mu^-$  decays in the Standard Model and beyond, JHEP **0901** (2009) 019, arXiv:0811.1214.
- [18] M. Neubert, Heavy quark symmetry, Phys. Rept. 245 (1994) 259, arXiv:hep-ph/9306320.
- [19] V. Belyaev, A. Khodjamirian, and R. Ruckl, QCD calculation of the  $B \to \pi, K$  form-factors, Z. Phys. C60 (1993) 349, arXiv:hep-ph/9305348.
- [20] P. Colangelo and A. Khodjamirian, QCD sum rules, a modern perspective, arXiv:hep-ph/0010175.
- [21] V. M. Braun, Light cone sum rules, arXiv:hep-ph/9801222.
- [22] A. Khodjamirian and R. Ruckl, QCD sum rules for exclusive decays of heavy mesons, Adv. Ser. Direct. High Energy Phys. 15 (1998) 345, arXiv:hep-ph/9801443.
- [23] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, Comparative study of the decays B → (K, K\*)ℓ<sup>+</sup>ℓ<sup>-</sup> in standard model and supersymmetric theories, Phys. Rev. D61 (2000) 074024, arXiv:hep-ph/9910221.
- [24] J. J. Wang, R. M. Wang, Y. G. Xu, and Y. D. Yang, The rare decays B<sup>+</sup> → π<sup>+</sup>ℓ<sup>+</sup>ℓ<sup>-</sup>, ρ<sup>+</sup>ℓ<sup>+</sup>ℓ<sup>-</sup>, B<sup>0</sup> → μ<sup>+</sup>μ<sup>-</sup> in the R-parity violating supersymmetry, Phys. Rev. D77 (2008) 014017, arXiv:0711.0321.
- [25] S. Fajfer, J. F. Kamenik, and I. Nisandzic, On the  $B \to D^* \tau \bar{\nu}_{\tau}$  Sensitivity to New Physics, Phys. Rev. **D85** (2012) 094025, arXiv:1203.2654.
- [26] BaBar Collaboration, J. Lees et al., Measurement of an Excess of  $B \to D^{(*)} \tau \nu$ decays and implications for charged Higgs bosons, arXiv:1303.0571.
- [27] LHCb collaboration, A. A. Alves Jr et al., The LHCb detector at the LHC, Journal of Instrumentation 3 (2008), no. 08 S08005.

- [28] LHCb Collaboration, R. Antunes-Nobrega et al., LHCb reoptimized detector design and performance: Technical Design Report, Technical Design Report LHCb, CERN, Geneva, 2003.
- [29] LHCb Collaboration, P. R. Barbosa-Marinho et al., LHCb inner tracker: Technical Design Report, Technical Design Report LHCb, CERN, Geneva, 2002. revised version number 1 submitted on 2002-11-13 14:14:34.
- [30] J. Van Tilburg and M. Merk, Track simulation and reconstruction in LHCb, PhD thesis, Vrije U. Amsterdam, Amsterdam, 2005, Presented on 01 Sep 2005.
- [31] LHCb Collaboration, S. Amato et al., LHCb calorimeters: Technical Design Report, Technical Design Report LHCb, CERN, Geneva, 2000.
- [32] T. Blake, Ring Imaging Cherenkov Detectors and the Rare Decay  $B_d \to K^{*0} \mu^+ \mu^$ at LHCb, PhD thesis, London U., London, 2008.
- [33] LHCb RICH group, M. Adinolfi et al., Performance of the LHCb RICH detector at the LHC, Eur. Phys. J. C 73 (2012) 2431. 25 p.
- [34] LHCb Collaboration, P. R. Barbosa-Marinho et al., LHCb muon system: Technical Design Report, Technical Design Report LHCb, CERN, Geneva, 2001.
- [35] S. Kullback and R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1951), no. 1 pp. 79.
- [36] M. Needham, Clone track identification using the Kullback-Liebler distance, Tech. Rep. LHCb-2008-002. CERN-LHCb-2008-002. LPHE-2008-002, CERN, Geneva, 2008.
- [37] T. Sjöstrand, S. Mrenna and P. Skands, PYTHIA 6.4 physics and manual, JHEP 05 (2006) 026, arXiv:hep-ph/0603175.
- [38] I. Belyaev et al., Handling of the generation of primary events in GAUSS, the LHCb simulation framework, Nuclear Science Symposium Conference Record (NSS/MIC) IEEE (2010) 1155.
- [39] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instrum. Meth. A462 (2001) 152.
- [40] P. Golonka and Z. Was, PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays, Eur. Phys. J. C45 (2006) 97, arXiv:hep-ph/0506026.
- [41] Geant4 collaboration, S. Agostinelli et al., Geant4: a simulation toolkit, Nucl. Instrum. Meth. A506 (2003) 250.

- [42] Geant4 collaboration, J. Allison et al., Geant4 developments and applications, IEEE Trans. Nucl. Sci. 53 (2006) 270.
- [43] The LHCb collaboration, The Boole project, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/.
- [44] A. Jaeger et al., Measurement of the track finding efficiency, Tech. Rep. LHCb-PUB-2011-025. CERN-LHCb-PUB-2011-025, CERN, Geneva, Apr, 2012.
- [45] M. Alexander, 2013. https://lbtwiki.cern.ch/bin/view/VELO/VELOConfPlots\_IPResols, accessed 11 February, 2014.
- [46] A. Powell, 2013. https://lbtwiki.cern.ch/bin/view/RICH/RichPIDPerformancePlots, accessed 11 February, 2014.
- [47] F. Archilli et al., Performance of the muon identification at LHCb, Journal of Instrumentation 8 (2013) 20P, arXiv:1306.0249.
- [48] R. Aaij et al., The LHCb trigger and its performance in 2011, Journal of Instrumentation 8 (2013), no. 04 P04022, arXiv:1211.3055.
- [49] R. E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME Journal of Basic Engineering (1960), no. 82 (Series D) 35.
- [50] R. Fruhwirth, Application of Kalman filtering to track and vertex fitting, Nucl. Instrum. Meth. A262 (1987) 444.
- [51] SLD Collaboration, K. Abe et al., Measurement of the b quark fragmentation function in Z<sup>0</sup> decays, Phys. Rev. D65 (2002) 092006, arXiv:hep-ex/0202031.
- [52] V. V. Gligorov and M. Williams, Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree, Journal of Instrumentation 8 (2013) 2013P, arXiv:1210.6861.
- [53] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.
- [54] R. E. Schapire and Y. Freund, A decision-theoretic generalization of on-line learning and an application to boosting, Jour. Comp. and Syst. Sc. 55 (1997) 119.
- [55] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss, The Toolkit for MultiVariate data Analysis, TMVA 4, J. Phys. Conf. Ser. 219 (2010) 032057.

- [56] S. M. Bilenky, S. Pascoli, and S. T. Petcov, Majorana neutrinos, neutrino mass spectrum, CP-violation and neutrinoless double beta-decay. I: The three-neutrino mixing case, Phys. Rev. D64 (2001) 053010, arXiv:hep-ph/0102265.
- [57] R. N. Mohapatra and A. Y. Smirnov, Neutrino mass and new physics, Ann. Rev. Nucl. Part. Sci. 56 (2006) 569, arXiv:hep-ph/0603118.
- [58] A. Atre, T. Han, S. Pascoli, and B. Zhang, The Search for heavy majorana neutrinos, JHEP 05 (2009) 030, arXiv:0901.3589.
- [59] S. M. Bilenky, Neutrinoless double beta-decay, Phys. Part. Nucl. 41 (2010) 690, arXiv:1001.1946.
- [60] CLEO, K. W. Edwards et al., Search for lepton flavor violating decays of B mesons, Phys. Rev. D65 (2002) 111102, arXiv:hep-ex/0204017.
- [61] Particle Data Group, K. Nakamura et al., Review of particle physics, J. Phys. G37 (2010) 075021.
- [62] T. Skwarnicki, A study of the radiative cascade transitions between the Upsilonprime and Upsilon resonances, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [63] LHCb Collaboration, R. Aaij *et al.*, Search for Majorana neutrinos in  $B^- \to \pi^- \mu^+ \mu^+$  decays, Phys. Rev. Lett. **112** (2014) 131802.
- [64] W. Bonivento *et al.*, Proposal to search for heavy neutral leptons at the SPS, arXiv:1310.1762.
- [65] LHCb collaboration, R. Aaij *et al.*, Measurement of the  $B_s^0 \overline{B}_s^0$  oscillation frequency  $\Delta m_s$  in  $B_s^0 \to D_s(3)\pi$  decays, Phys. Lett. **B709** (2012) 177, arXiv:1112.4311.
- [66] CDF collaboration, A. Abulencia *et al.*, Observation of  $B_s^0 \overline{B}_s^0$  oscillations, Phys. Rev. Lett. **97** (2006) 242003.
- [67] A. Bazavov et al., Neutral B-meson mixing from three-flavor lattice QCD: determination of the SU(3)-breaking ratio  $\xi$ , arXiv:1205.7013.
- [68] Heavy Flavor Averaging Group, Y. Amhis et al., Averages of b-hadron, c-hadron, and τ-lepton properties as of early 2012, arXiv:1207.1158.
- [69] BaBar collaboration, P. del Amo Sanchez *et al.*, Study of  $B \to X\gamma$  decays and determination of  $|V_{td}/V_{ts}|$ , Phys. Rev. **D82** (2010) 051101, arXiv:1005.4087.
- [70] Belle collaboration, K. Abe *et al.*, Observation of  $b \to d\gamma$  and determination of  $|V_{td}/V_{ts}|$ , Phys. Rev. Lett. **96** (2006) 221601, arXiv:hep-ex/0506079.

- [71] BaBar collaboration, B. Aubert *et al.*, Branching fraction measurements of  $B^+ \rightarrow \rho^+ \gamma$ ,  $B^0 \rightarrow \rho^0 \gamma$ , and  $B^0 \rightarrow \omega \gamma$ , Phys. Rev. Lett. **98** (2007) 151802, arXiv:hep-ex/0612017.
- [72] T. Hurth and M. Nakao, Radiative and electroweak penguin decays of B mesons, Ann. Rev. Nucl. Part. Sci. 60 (2010) 645, arXiv:1005.1224.
- [73] A. Buras et al., Universal unitarity triangle and physics beyond the standard model, Phys. Lett. B500 (2001) 161, arXiv:hep-ph/0007085.
- [74] T. Feldmann and T. Mannel, Minimal flavour violation and beyond, JHEP 02 (2007) 067, arXiv:hep-ph/0611095.
- [75] R. Barbieri, D. Buttazzo, F. Sala, and D. M. Straub, Less minimal flavour violation, arXiv:1206.1327.
- [76] Belle collaboration, J. Wei *et al.*, Search for  $B \to \pi \ell^+ \ell^-$  decays at Belle, Phys. Rev. **D78** (2008) 011101, arXiv:0804.3656.
- [77] LHCb collaboration, R. Aaij et al., Differential branching fraction and angular analysis of the  $B^+ \to K^+ \mu^+ \mu^-$  decay, arXiv:1209.4284.
- [78] T. Skwarnicki, A study of the radiative cascade transitions between the Upsilonprime and Upsilon resonances, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [79] P. Ball and R. Zwicky, New results on  $B \to \pi, K, \eta$  decay form factors from lightcone sum rules, Phys. Rev. **D71** (2005) 014015, arXiv:hep-ph/0406232.
- [80] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instrum. Meth. A462 (2001) 152.
- [81] A. Ali, E. Lunghi, C. Greub, and G. Hiller, Improved model independent analysis of semileptonic and radiative rare B decays, Phys. Rev. D66 (2002) 034002, arXiv:hep-ph/0112300.
- [82] H. Asatryan, H. Asatrian, C. Greub, and M. Walker, Two loop virtual corrections to  $B \rightarrow X_s \ell^+ \ell^-$  in the standard model, Phys. Lett. **B507** (2001) 162, arXiv:hep-ph/0103087.
- [83] C. Bobeth, M. Misiak, and J. Urban, Photonic penguins at two loops and m(t) dependence of  $BR[B \to X_s \ell^+ \ell^-]$ , Nucl. Phys. **B574** (2000) 291, arXiv:hep-ph/9910220.
- [84] Charged Higgs effects on exclusive semi-tauonic B decays, Zeitschrift fr Physik C Particles and Fields 67 (1995), no. 2.

- [85] Belle Collaboration, A. Matyja *et al.*, Observation of  $B^0 \to D^{*-} \tau \nu_{\tau}$  decay at Belle, Phys. Rev. Lett. **99** (2007) 191807, arXiv:0706.4429.
- [86] Belle Collaboration, I. Adachi *et al.*, Measurement of  $B \to D^{(*)} \tau \nu$  using full reconstruction tags, arXiv:0910.4301.
- [87] Belle Collaboration, A. Bozek *et al.*, Observation of  $B^+ \to \overline{D}^{*0} \tau^+ \nu_{\tau}$  and Evidence for  $B^+ \to \overline{D}^0 \tau^+ \nu_{\tau}$  at Belle, Phys. Rev. **D82** (2010) 072005, arXiv:1005.2302.
- [88] A. Crivellin, C. Greub, and A. Kokulu, Explaining  $B \to D\tau\nu$ ,  $B \to D^*\tau\nu$  and  $B \to \tau\nu$  in a 2HDM of type III, Phys. Rev. **D86** (2012) 054014, arXiv:1206.2634.
- [89] P. Biancofiore, P. Colangelo, and F. De Fazio, On the anomalous enhancement observed in  $B \to D^{(*)}\tau\bar{\nu}_{\tau}$  decays, Phys. Rev. **D87** (2013), no. 7 074010, arXiv:1302.1042.
- [90] Y. Sakaki, M. Tanaka, A. Tayduganov, and R. Watanabe, *Testing leptoquark models*  $in \ \bar{B} \to D^{(*)}\tau\bar{\nu}$ , Phys. Rev. **D88** (2013) 094012, arXiv:1309.0301.
- [91] I. Dorner, S. Fajfer, N. Konik, and I. Niandi, Minimally flavored colored scalar in  $\bar{B} \rightarrow D^{(*)}\tau\bar{\nu}$  and the mass matrices constraints, JHEP **1311** (2013) 084, arXiv:1306.6493.
- [92] W. Altmannshofer, M. Carena, N. R. Shah, and F. Yu, *Indirect probes of the MSSM after the Higgs discovery*, JHEP **1301** (2013) 160, arXiv:1211.1976.
- [93] U. Haisch and F. Mahmoudi, MSSM: cornered and correlated, JHEP 1301 (2013) 061, arXiv:1210.7806.
- [94] LHCb collaboration, R. Aaij et al., Study of D<sub>J</sub> meson decays to D<sup>+</sup>π<sup>-</sup>, D<sup>0</sup>π<sup>+</sup> and D<sup>\*+</sup>π<sup>-</sup> final states in pp collisions, J. High Energy Phys. **09** (2013) 145. 30 p, Comments: 30 pages, 12 pdf figures, submitted to JHEP.
- [95] S. Godfrey and N. Isgur, Mesons in a relativized quark model with chromodynamics, Phys. Rev. D32 (1985) 189.
- [96] F. U. Bernlochner, Z. Ligeti, and S. Turczyk, A Proposal to solve some puzzles in semileptonic B decays, Phys. Rev. D85 (2012) 094033, arXiv:1202.1834.
- [97] ALEPH Collaboration, R. Barate et al., Observation of doubly charmed B decays at LEP, Eur. Phys. J. C4 (1998) 387.
- [98] LHCb collaboration, R. Aaij *et al.*, Measurement of the  $B_c^+$  meson lifetime using  $B_c^+ \to J/\psi \mu^+ \nu_\mu X$  decays, .

- [99] CDF Collaboration, A. Abulencia *et al.*, Search for  $B_s^0 \to \mu^+\mu^-$  and  $B_d^0 \to \mu^+\mu^$ decays in  $p\overline{p}$  collisions with CDF II, Phys. Rev. Lett. **95** (2005) 221805.
- [100] D0 Collaboration, V. M. Abazov *et al.*, Search for the rare decay  $B_s^0 \to \mu^+\mu^-$ , Phys. Rev. D 87 (2013) 072006.
- [101] CMS Collaboration, S. Chatrchyan et al., Measurement of the B<sub>s</sub> → μ<sup>+</sup>μ<sup>-</sup> branching fraction and search for B<sup>0</sup> → μ<sup>+</sup>μ<sup>-</sup> with the CMS Experiment, Phys. Rev. Lett. **111** (2013) 101804, arXiv:1307.5025.
- [102] M. Pivk and F. R. Le Diberder, Plots: a statistical tool to unfold data distributions, Nuclear Instruments and Methods in Physics Research A 555 (2005) 356, arXiv:physics/0402083.
- [103] D. Scora and N. Isgur, Semileptonic meson decays in the quark model: An update, Phys. Rev. D52 (1995) 2783, arXiv:hep-ph/9503486.
- [104] E. Bulbul et al., Detection of an unidentified emission line in the stacked X-ray spectrum of galaxy clusters, arXiv:1402.2301.
- [105] A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, and J. Franse, An unidentified line in X-ray spectra of the Andromeda galaxy and Perseus galaxy cluster, arXiv:1402.4119.
- [106] LHCb collaboration, R. Aaij et al., Measurement of form-factor independent observables in the decay  $B^0 \rightarrow K^{*0}\mu^+\mu^-$ , Phys. Rev. Lett. **111** (2013) 191801, arXiv:1308.1707.
- [107] J. Matias, F. Mescia, M. Ramon and J. Virto, Complete anatomy of  $B^0 \to K^{*0} \mu^+ \mu^$ and its angular distribution, JHEP **1204** (2012) 104, arXiv:1202.4266.
- [108] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, Optimizing the basis of  $B \rightarrow K^* \ell^+ \ell^-$  observables in the full kinematic range, JHEP **1305** (2013) 137, arXiv:1303.5794.
- [109] S. Jäger and J. Martin Camalich, On  $B \rightarrow V l l$  at small dilepton invariant mass, power corrections, and new physics, JHEP **1305** (2013) 043, arXiv:1212.2263.
- [110] W. Altmannshofer and D. M. Straub, New physics in  $B \to K^* \mu \mu$ ?, Eur. Phys. J. C73 (2013) 2646, arXiv:1308.1501.
- [111] S. Descotes-Genon, J. Matias, and J. Virto, Understanding the  $B \to K^* \mu^+ \mu^$ anomaly, Phys. Rev. **D88** (2013) 074002, arXiv:1307.5683.

- [112] F. Beaujean, C. Bobeth, and D. van Dyk, Comprehensive Bayesian Analysis of Rare (Semi)leptonic and Radiative B Decays, arXiv:1310.2478.
- [113] C. Bouchard *et al.*, Standard Model predictions for  $B \to Kll$  with form factors from lattice QCD, Phys. Rev. Lett. **111** (2013) 162002, arXiv:1306.0434.
- [114] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, Calculation of  $B^0 \to K^{*0}\mu^+\mu^-$  and  $B_s^0 \to \phi\mu^+\mu^-$  observables using form factors from lattice QCD, arXiv:1310.3887.
- [115] K. Petridis, Introduction to heavy flavours, Presented at Moriond QCD and High Energy Interactions.

# Glossary

SM	Standard Model
СР	charge-parity
u MSM	Minimal Neutrino Standard Model
QCD	Quantum ChromoDynamics
LHC	Large Hadron Collider
СКМ	Cabibbo-Kobayashi-Maskawa
OPE	Operator Product Expansion
MFV	Minimal Flavour Violation
HQET	Heavy Quark Effective Theory
LCSR	Light Cone Sum Rules
PV	Primary Vertex
SV	Secondary Vertex
VELO	VErtex LOcator
тт	Tracker Turicensis
RICH	Ring Imaging CHerenkov
ІТ	Inner Tracker
от	Outer Tracker
ST	Silicon Tracker
PID	Particle IDentification
DLL	Difference in Log-Likelihood
ECAL	Electromagnetic CALorimeter
HCAL	Hadronic CALorimeter
SPD	Scintillator Pad Detector

PS	Pre-Shower
MWPC	Multi-Wire Proportion Chamber
GEM	Gaseous Electron Multiplier
isMuon	Standard muon identification criteria
isMuonLoose	Looser muon identification criteria
IP	Impact Parameter
KL	Kullback-Leibler
LO	First Level Trigger
HLT	High Level Trigger
HLT1	First stage of the HLT
HLT2	Second stage of the HLT
DOCA	Distance Of Closest Approach
MVA	MultiVariate Analysis
BDT	Boosted Decision Tree
PDF	Probability Density function
2HDM	Two Higgs Doublet Model
MSSM	Minimal Supersymmetric Standard Model

# Appendix A

# $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ data-simulation comparisons



Figure A.1: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for the  $B^+ \chi^2_{IP}$ . The smeared (red) and unsmeared (green) simulated events are shown against data (black).



Figure A.2: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for relevant  $B^+$  variables in  $B^+ \to J/\psi K^+$ .



Figure A.3: Comparisons between data and simulated  $B^+ \rightarrow J/\psi K^+$  events for relevant  $K^+$  variables in  $B^+ \rightarrow J/\psi K^+$ .


Figure A.4: Comparisons between data and simulated  $B^+ \rightarrow J/\psi K^+$  events for relevant  $\mu^+$  variables in  $B^+ \rightarrow J/\psi K^+$ .



Figure A.5: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for relevant  $\mu^-$  variables in  $B^+ \to J/\psi K^+$ .



Figure A.6: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for  $\Delta P(\mu^+\mu^-)$  in  $B^+ \to J/\psi K^+$ .



Figure A.7: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for the  $B^+ \chi^2_{IP}$ , after reweighting.



Figure A.8: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for relevant  $B^+$  variables in  $B^+ \to J/\psi K^+$ , after reweighting.



Figure A.9: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for relevant  $K^+$  variables in  $B^+ \to J/\psi K^+$ , after reweighting.



Figure A.10: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for relevant  $\mu^+$  variables in  $B^+ \to J/\psi K^+$ , after reweighting.



Figure A.11: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for relevant  $\mu^-$  variables in  $B^+ \to J/\psi K^+$ , after reweighting.



Figure A.12: Comparisons between data and simulated  $B^+ \to J/\psi K^+$  events for  $\Delta P(\mu^+\mu^-)$  in  $B^+ \to J/\psi K^+$ , after reweighting.